



SEQUENCE LISTING

<110> MBARI

DeLong, Edward  
Beja, Oded

<120> Light-driven energy generation using proteorhodopsin

<130> MBA-101

<140> US/09/847,513

<141> 2001-05-01

<150> 60/201,602

<151> 2000-05-03

<160> 65

<170> PatentIn version 3.0

<210> 1

<211> 105184

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> gene

<222> (50866)..(51615)

<223> Proteorhodopsin gene sequence.

<220>

<221> misc\_feature

<222> (1593)..(2807)

<223> Predicted threonine dehydratase. Contains 'n' at position 2753.

<300> Beja, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P.,  
 Jovanovich, S.B., Gates, C.M., Feldman, R.A., DeLong, E.F  
 <302> Bacterial rhodopsin: evidence for a new type of phototrophy in the sea  
 <303> Science  
 <304> 289  
 <305> 5486  
 <306> 1902-1906  
 <307> 2000-09-15  
 <308> AF279106  
 <309> 2000-06-15  
 <313> (50866)..(51615)

```

<400> 1
ttgttatatc agtaatggct attgctccaa taacttaata ctaatatata attagtttat    60

gaataaaattt tatatatattg ggttattgtt ttttacacta aatgcatttt cttgctcaga    120

tcttctagat acagacatga gagttcttga ttccgctgag tcaagaaaacc tttgcgagtt    180

tgaaggaaaa gcttttactag ttgtgaatgt tgcaagtaga tgtgggttaca cttatcaata    240

tgctggcctt caaaagtatt atgaaagtta taaagatgaa gattttctag taattgggat    300

cccatctaga gatttttcttc aagaatactc tgatgaaagc gatgttgcag aattttgttc    360

tacagaatac ggtgttgaat ttcttatgtt ctcaactgct aaagtcaaag gaaaaaaagc    420

acacccattt tataaaaaac ttattgcaga atcagggtttt actccctcat ggaactttaa    480

taaataactta atctcaaaaag agggcaaggt tgtatccaca tatggatcaa aggtaaagcc    540
  
```

|             |             |              |             |             |             |      |
|-------------|-------------|--------------|-------------|-------------|-------------|------|
| tgattcaaaa  | gagcttatat  | cagctataga   | aggcttgctg  | taaaattatt  | acttagaaac  | 600  |
| taatacagtt  | ttaggcttgt  | ttgctgcaa    | tattccatta  | tctacaactc  | caggaatatt  | 660  |
| attaatcaaa  | gcttccattt  | cagtgggggt   | tgaaatatcc  | atattagaga  | tatctaaaaa  | 720  |
| gtgattacct  | tggtctgtta  | taaatccagt   | tctatatgtg  | ggtattccac  | cgatcgagat  | 780  |
| tatttttctt  | gcaacaaggc  | tcctactttc   | aggtatcacc  | tctataggca  | gtggaaaaagc | 840  |
| tcccaaaaaga | ttaaccatct  | ttgactgata   | aactatacat  | ataaaactgt  | tagaggcaga  | 900  |
| agcaactatc  | ttttctctag  | tatgtgcgcc   | accaccacct  | ttaataagac  | aattttcagg  | 960  |
| agacacctca  | tctgcaccat  | ctatgtaata   | agctatatca  | actacatcat  | taaggctaaa  | 1020 |
| gacctctatc  | ccatttttcat | ttaataaattt  | tgatgaagca  | tctgaactag  | aaacagctcc  | 1080 |
| agcaaaatttg | tgccctatgct | ccttttagttc  | ttctataaaa  | aaattaactg  | ttgagccgggt | 1140 |
| tccaataacct | aaaatcatct  | caggatgaag   | attattttttg | atataattcta | tagcttgttt  | 1200 |
| agcaacattt  | atctttgagc  | cactcataga   | gttataatac  | aagaaaaatat | aggtagttaa  | 1260 |
| ttatttttgag | actaaaaaatt | aaaaaaaaacag | gttctttttaa | gaattcccag  | aagtacctaa  | 1320 |
| agcttattcc  | taatgcattc  | gtttatgacg   | tagcattaaa  | atcacctata  | acatttgctc  | 1380 |
| taaatatattc | ttcaaagctg  | gggaataaaa   | ttttcctaata | aagagaggat  | ctgcaaaccta | 1440 |
| tatttttctt  | taaaaaacaga | ggagcgtata   | acaagattgt  | aaatttatcc  | gatgccgaaa  | 1500 |
| agaagagggg  | ggttatttgc  | gcatacagcag  | gaaatcatgc  | tcaaggggta  | gccagtgcat  | 1560 |
| gtaagaaaatt | aaaaaattaat | tgcttgatag   | ttatgccaat  | aacaactcca  | gaaataaaaa  | 1620 |

|   |      |
|---|------|
| taaaagatgt aaaaagattt ggagccaaaa tactccaaca tggggacaac gtagatgcag   | 1680 |
| cattaaaga ggcactgttt attgcaaaga aaaaaaaaaatt gtctttttgtt catccttttg | 1740 |
| acgacctctt aacaattgct ggccaaggga ctataggaca agaaattctt gaagataaaa   | 1800 |
| ataattttga tgttgtcttt gtccggtgg gaggaggagg tattctagct ggtgtatctg    | 1860 |
| cctggatagc acagaataat aagaaaaataa aaattgttgg tgttgagggtt gaggattccg | 1920 |
| cttgtcttgc tgaggccgta aaagctaata aaagagttat tttaaaagaa gtgggcctct   | 1980 |
| ttgctgatgg ggtggcagta tcaagggttg gaaaaataaa ttttgatgtt attaaagagt   | 2040 |
| gcgtagatga agtcattaca gttagcgttg atgaggcttg caccgctgta aaagatatct   | 2100 |
| ttgaagatac aagggttcta tcagaacctg ctggggcatt agcacttgca gggttaaaaag  | 2160 |
| cctacgcaag gaaagttaaa aataaaaaaac ttattgctat aagttctggc gctaattgtaa | 2220 |
| atttccaaag acttaatttt attgttgagc gatcagagat tggtgaaaaat agagaaaaaa  | 2280 |
| tattaaagtat caaaatccca gagataacctg gaagttttct taagctttca aggatgtttg | 2340 |
| gcagctctca agttacagag tttaactaca ggaaatctag cttaagcgat gcatatgttt   | 2400 |
| tagttggtgt tagaaactaaa actgaaaaat catttgaaat cttaaagtcc aaattaaaaa  | 2460 |
| aagcaggctt caccttttagc gactttactc gaaatgaaat atccaatgat catctgaggc  | 2520 |
| atatggttgg tggcagaaat agtgactcag gctctcataa caatgaaaga atatttaggg   | 2580 |
| gagagtttcc tgagaagccg ggcgcgctgt taaattttct agagaaaattt ggaaataaat  | 2640 |

|            |             |             |             |             |              |      |
|------------|-------------|-------------|-------------|-------------|--------------|------|
| ggmatatttc | cttattttcat | tacaggaacc  | taggttcagc  | ttttggaaag  | atattaattg   | 2700 |
| gcatcgagag | taaggataaa  | gacaagctaa  | taaatcattt  | aaataagtca  | ggnactattt   | 2760 |
| ttacagaaga | aacctctaac  | aaggcataca  | aagatttttt  | aaaatgaaag  | gttaataactt  | 2820 |
| taatctaaat | ttaattgaaa  | aaagctcatc  | gctagggttt  | tcccacggct  | ctttgaacaa   | 2880 |
| ctcggattga | gatctatcat  | cctcctcgtc  | gtaaattctc  | ccaccctttag | aatagaccaa   | 2940 |
| aaatagatat | gacaaaggag  | cgagctcata  | tttatatcta  | atttggaacg  | aagctacgcc   | 3000 |
| agtattaaat | tcattaacta  | tattatttcc  | tttataaagg  | tatccattag  | catctgaaaa   | 3060 |
| aatactaata | ggatttttttg | ctttcaaagc  | aacaaaattga | ctcttaagtc  | tgatctcatg   | 3120 |
| tttattattt | ttaaaccaat  | ttagatcaaa  | agataaggta  | tcttgtcttg  | aatcatatatga | 3180 |
| ggcaagatta | ttattatcct  | gccatatcag  | ccatttcattt | tcctttctta  | ttctatatattg | 3240 |
| cgcattgatt | cttaagttat  | catttggaaa  | tattgaaacct | gctatcttgt  | aaaaattttct  | 3300 |
| accataccca | ttcgaatccc  | accgattatc  | tttctctccc  | ttaaagaagc  | taactctcca   | 3360 |
| gtcatatgtc | cagaatgaat  | agttcttttg  | ctcaaagtct  | gctgtaatac  | ctattcttctt  | 3420 |
| ttttgatttg | ataaaggggt  | aggcttcatt  | ttttcttgtg  | atagttgtat  | ttttcccaga   | 3480 |
| agatctaaag | ttaaaaatcta | attgaaaattt | agagttgtcc  | ttaaaaactaa | aagaatttttt  | 3540 |
| ttgatcgatg | cctattggat  | ttgaattacc  | agctgtgtca  | gcatacataat | ttagatcaat   | 3600 |
| tccataatct | atttggttta  | atatcgagct  | attatcaaat  | tcattttattt | ttcgattttcc  | 3660 |
| tccaatacca | gcataaatcc  | agtctcttct  | ttgcagataa  | ccaaagtcat  | ttaactcgaa   | 3720 |

|            |             |            |             |            |             |      |
|------------|-------------|------------|-------------|------------|-------------|------|
| gtcgtcttca | aaataaagaa  | ggcttccact | tatgttagat  | agttatttg  | gaagatatgt  | 3780 |
| aaactgagtc | ctatacccaa  | gccattttt  | accatctttt  | tctgaggcta | acagatctga  | 3840 |
| atatgtaatt | aatttttttg  | aacgaatatt | gatgtaatca  | ataacattga | ctgttgatga  | 3900 |
| ttcgcctgtc | atttcattct  | caacattcgt | taccatgaag  | ccaagcgttt | tatttccaag  | 3960 |
| ctttgttcga | gatcgaagag  | cataatagtc | tcttccaact  | gaaaaggctt | catcagcctc  | 4020 |
| acttgctaca | aatactccaa  | attcattatt | attacttttt  | tgagtaagtc | ttaatgcaaa  | 4080 |
| atcaatatca | gaatagtttt  | ttttagctgc | ctcgcaacct  | tcttcattac | tttcttctga  | 4140 |
| gcaattatag | ctgggggcag  | ctccaatcct | cgcgtatatt  | ataaccgagt | atctatcata  | 4200 |
| attactaata | tcaaatagtg  | attggttttc | attgaaaaat  | gctctttttt | ctgagtaaaa  | 4260 |
| agtttcttga | gcagaaaaagt | taataaccac | atcatcactc  | tcagcttgtc | cgaaatctgg  | 4320 |
| attaatagct | aaatttat    | gacgaccttt | tccagtgcta  | taaaagattt | cagccccaat  | 4380 |
| atctgaccct | tcttggtttg  | taactgaatt | tttatattgaa | gatatatatg | gaaaaaaaagt | 4440 |
| aagctttgat | tttgtatagt  | tttgtatttc | taagctatct  | aactcttgaa | agtagtcatt  | 4500 |
| tctactagcc | attgttcctg  | cactgctaac | ccatgactca  | ttttcggcac | tataacgtaa  | 4560 |
| tgccgtgtaa | ttaatttttc  | ttatatcacc | atcaggctgt  | ttcattaaag | ttacatccca  | 4620 |
| aggaataaaa | aactcagaga  | cccaataccc | atcaaat     | ttgttttttg | caatccaatc  | 4680 |
| tccatccccc | tctgtctctaa | agtctcctgc | ttgcgttttt  | atggcatcga | aaagcgagtt  | 4740 |

|  |      |
|--|------|
| ccaagattt atagcaagaa tgaaagcttt gttaccatca ccatcaaagt ctatatattat    | 4800 |
| agagttttta tcgctaagtg aatttatattg atctctaagc gtcttcctgg agaacataga   | 4860 |
| atcattactt tgaaaaattct taaatccaac atatatacca tccttatttg agaaaaattaa  | 4920 |
| agccgttgta agaagttcat ttttcttaag agtaaaagga gatgtttcat aaaaatctgt    | 4980 |
| aatttcaaat gcattattcc actcaggctc atcaagagag ccatcaataa caattgagtt    | 5040 |
| tgaccaaatc agcacagagg taagtaaaaag tgataatgag gctaaaagtta ttttcataga  | 5100 |
| tagatttaat tcagagtatt ttaatctatg aagagtagga aatcatccga tcattctcag    | 5160 |
| aaaaaacata atcatattta aggtcatggt tttctccaaa atgcatatata caaatctggt   | 5220 |
| attgataaca cagtccaacc aataaaggtc ttgatacctc ctcgttata gagccaataa     | 5280 |
| ttctatcgaa atatccagag ccatagccaa gcctgtatcc atttaaatca actcctgtca    | 5340 |
| taggaataaa cattaaatca atttcattta tgttgacata atcttcaact ttaacctctt    | 5400 |
| tgatcccaaa ttgatttata aagaagttag gctgctcatc cagcaaaatta aaagccatca   | 5460 |
| tttcgtcatc aattactttt ggaatataaaa tattctttttt aagtttagta aaagcttgaa  | 5520 |
| ttaacaaaatt tgtattgact tcatttcgaa aaggaaaaata taaggcaatg ttttgcatTT  | 5580 |
| catgagtatt gatTTTTTct aatacatTTTt cttggattaa aaaacttata ttgctctcttag | 5640 |
| ataaatcaga aatagactgc ccctgctcaa agagtgattt tcgtatttta tttttcacca    | 5700 |
| taaattgtgc cgaggctaac aaaccaaatac accgccatgt gctataacctg aaccgatggg  | 5760 |
| tcagggtggtg aatttcataa caaatcaggc ttccctgtaa aggtactgcg caacaatgct   | 5820 |

|             |             |             |            |            |            |      |
|-------------|-------------|-------------|------------|------------|------------|------|
| agaggaatat  | cacctattta  | attgtatcgg  | ttcaaattha | ttaacacatt | agcgtatgaa | 5880 |
| ccagaatgca  | gttaattata  | aaatatatag  | gtattaagta | aaagttaatt | tttagagagc | 5940 |
| agactctatt  | ttttgtatta  | gcttttcgat  | atctttatta | tttaaagcag | agtcagtatt | 6000 |
| ctcggttggc  | ttggataaaa  | gcttactagc  | catagttaga | ccagctatca | ccaaggcatt | 6060 |
| attcttatca  | ctgatgccat  | caagctcatc  | atttaataat | tttgcagctc | taatgagttg | 6120 |
| gtctttctct  | tctggcggac  | aggctaact   | cagatctcgg | ccaaaaattc | ttaaagatag | 6180 |
| cgtttccatt  | tctgacatta  | tttttttgagt | ttagttattt | ctttatttag | tatgagctta | 6240 |
| tcttttttcc  | agtcctcttc  | cctttgcaca  | taagaatcta | ttatccctt  | ttgctcatca | 6300 |
| aatcctttta  | tcaaaagatc  | taccttatcc  | tctagttcaa | ataatgatga | ttttttttct | 6360 |
| tctgacattg  | tataagttta  | attgtactag  | aatgaattgg | aaagtttctt | taataaaaga | 6420 |
| gtaaaataag  | gcatggaaaa  | aataattttt  | aaaaacagaa | gagattcact | tataaaacac | 6480 |
| ctacctaaaga | atcagcctt   | aattgtacct  | ggtgcagatt | tgcaatatag | aaatgctgac | 6540 |
| tcactctata  | atttaagaca  | ggaaagtagc  | ttctactatc | tgtctggctt | ttgtgagccc | 6600 |
| tcttctctaa  | tggtttttagt | taataatgga  | aaaagcattg | attcaataat | ttttgttctt | 6660 |
| gaaaaagata  | aacttaaaga  | aatctgggat  | ggttatcgag | ctggccctga | gggcgcaata | 6720 |
| aatgattttc  | tttttgatca  | agcttttgaa  | aataataaat | cagatgcttt | aatgcctgaa | 6780 |
| atccttcaag  | ggctagaaaa  | agtttttttat | tcaataggga | agaaaaatgg | ctttgatcag | 6840 |



|  |      |
|--|------|
| aaagtaattg actggacatg cgcagcaaat tctaaagata ggcacagcaa atcaattgat  | 6900 |
| attattgatg gctcttcgat ggtaggaaat ttaaggctta tcaaagataa gcatgaaatt  | 6960 |
| gatatattga agagagcttg tgaaatttca gctgaatcat atattgaggt catgaaatct  | 7020 |
| ataaagcctg gggacaatga gcaggaaata gaggcgtat tttatatga attcgccaaa    | 7080 |
| aggggaggaa ggtttccagc ttatacacct atagttgctg gaggtgaggg tgctttgtga  | 7140 |
| ttgcattata ttgaaaaatga taaagagtta gcttcatcag atttaatttt ggtagacgca | 7200 |
| ggatgtgaat acaaaatgta tgcattctgat atcacaagaa ccttcccagt aagtggaaaa | 7260 |
| ttttcagatg aacagctaca aatttataat attgtccaca aagccaatct tgctgcaatc  | 7320 |
| gatgctgtaa aaactggtaa tagcataatg gagccccaaa tggtttcaga aaaagtaatt  | 7380 |
| actgaaggtc ttgtagagtt gggatatcta tctggcgatg ttaatcagct tcataaaaaat | 7440 |
| ggtgcattca aggactttta tatgcataag gtgggacatt ggcttggact tgatgttcat  | 7500 |
| gatgttggtg actacatgga gggagatgag tttatgaagt ttaagccagg gatgataacc  | 7560 |
| acaatcgagc caggcatcta tatcagtagc gcaatggatg tagatgacaa atggaaaaggc | 7620 |
| atcggcataa gaatagagga cgacatcctt gtaacagatt caggcaatat taatctaaca  | 7680 |
| gagaagggtc catctaatcc tcaagaaata gaatcattga tggcttagac tatggagggtt | 7740 |
| ccaattgtta tttctggcgg aggataata ggtaattaca tttctcttag gcttgaaaaa   | 7800 |
| aataatatca aaaccgttat tgtcgaaaaa gctagtagtt tcaaagccct agataagggt  | 7860 |
| ataagaacag tcactctcaa tgagcattct atgcaaatgc taaaaaatat tggttatttc  | 7920 |

|   |      |
|---|------|
| ccatcaattg ctcaaatcaa cagcatcgac gtattagatg gtgagggtac aggcaaaatt   | 7980 |
| caattttctag caaaggacgt aggcagcgaa aacctttcat atgtaaccta tttcaatgaa  | 8040 |
| ttacaaaaac taattttctga tccatgtaaa gaaagaacct tatttaataa tgagattgat  | 8100 |
| tcagttcaga atcttaatac agaattctgat ccagagatca tgcttaaaaga tggcatgacc | 8160 |
| ataaaaaacga atctaattgc tggatgtgat ggaagaaaatt caaatattgc aaaaattgct | 8220 |
| tcacttacaa gcagcttcga tgactactta caaacagctt taaactttgt cgttgatatt   | 8280 |
| gataatgatt cacatggcaa agctcaccaa gtttttttctg aaaaaggaat atttgcactt  | 8340 |
| atgccactcc cagaaggcaa ggtgagatg aataaatgca cagtggtttg gtcaataaaaa   | 8400 |
| aatcaagttt tgggagatga gcctgtatct gagtttgtaa aaaataacat ttcttttttt   | 8460 |
| gaatcaaaagc ttaatgttag tctcagggtt aagtcagaaaa ttttaagttt taaattatcg | 8520 |
| aaccatcatt ttgaaaaacta tattagcgga cctattgttc ttcttggtga tgctgctcac  | 8580 |
| tcaattcacc ccttagcagg tcaaggtatt aatctaggat ttgcagatgc agatactttt   | 8640 |
| tgtgaagagg taattagttc ttataaaaaa gggattgcct ttaatgagaa atcagtttta   | 8700 |
| aaaagatatg agattagaag aaaaagtatg aacttttttaa tgttgaagtc tatggacttt  | 8760 |
| tttgtggatt tatttggttc agaaaaattta tatctaaggc tgatcagaaa tttaggtatt  | 8820 |
| tcttcgttaa ataaatcaaa gtttgtttaa gcatttcttta taagacatgc ctctggaatg  | 8880 |
| aataagtttt aaatttgtat taaacttttt gaccttttagc tctaagttct ttaagaactt  | 8940 |

cactaatgcc ttttttatca atgattctca taccttttgc agatacttta agattttacga 9000  
acctgttctc agattcaacc caaaatttgt gtgtgtgaag attaggaaaa aactttcttt 9060  
tagtcctatt tttagcgtga gaaacattgt ttcctgactg tggatatctta cctgttactt 9120  
gacataatttt actcattgaa acgcgatttt atagaacact gaggaactta gcaatactat 9180  
tgtgaaacaa atttatttat tacggcatgc acaatctgac tgggagagct ctaatcagaa 9240  
agattttgat agaccattag caagaaaaagg cattgaagaa gcaataaaaa tatcatgtta 9300  
ctgcaaatct cattcaattt tagtagataa aatattctgt agcactgcag aaagaactaa 9360  
gcagactttt gatatatgca gtgatgggct taattatcca atagctgaag cagtctatac 9420  
tgatgagctt tacttttctg gccctgggtga gatcgttaag ctatccaaa gtttaagtga 9480  
attcatttcc tctgtttttaa taataggcca caatccatca atgcaaatgt acatagatgc 9540  
tatttcagaa aatcctcata ttacgtattc aacatgcggg ctggcagaaa ttctcgttga 9600  
aagttcatgg aaagacttat ctttaaaaa atgtaagtta aaatctttta ttcaaccagg 9660  
agagctttta aaattgaaaa acataaaaaat taagatcatt aaccactaa tgggatccaa 9720  
gataccctta cctcaatatg aaacaaaagg ctcggcagga ttggatttaa gggcatgcct 9780  
agatagtaat ctcagccttc aagcaggaac atctcagttg atacctattg gttttgcaat 9840  
gtacttagaa gatcccggtc ttgcagcaat ggtataacct agatcagggt taggttctaa 9900  
gcatggaatc gttcttggtga atctgggttg gttgattgat tcagactatc aaggagagct 9960  
aatggttcct gcctggaata gatcagatac agatttttgag attaatcctg gagacaggat 10020

|             |             |              |            |             |             |       |
|-------------|-------------|--------------|------------|-------------|-------------|-------|
| tgcacaaatg  | attatagttc  | cagtgattca   | agcagatttt | gaaattgtag  | acgagttcaa  | 10080 |
| tgagactcag  | aggggagaaa  | agggtttttgg  | aagttcaggt | ataaattgat  | aaattttactt | 10140 |
| tttcttgcca  | aatcttttctt | caaaattttctg | aactcttccg | ccagtatcaa  | taatttttttg | 10200 |
| cttaccggta  | taaaaaggat  | gagaagcaga   | ggatatatca | agagggtagt  | atgggtatgt  | 10260 |
| ttttccatct  | tcccattctt  | tcgttttgagt  | cgtatctaat | gttgaacgaa  | tgagaaaagaa | 10320 |
| cttatcagca  | ctagcgtcat  | ggaataaaac   | ttcacggtat | tcaggatgta  | tatctttttt  | 10380 |
| cataataaaa  | tcaaaaatttg | gatgagaact   | ataacaaaaa | acaacttctt  | ttcaatcaaa  | 10440 |
| acatgaaaat  | tttttactat  | gacatagctg   | tttcattgcc | tctaaggcaa  | tgcttttactt | 10500 |
| atagttctga  | gcttaaaaatt | acaaaaggaa   | cccgcgtgtc | agttcccttt  | ggtaaaaagga | 10560 |
| agattgtagg  | agtggtttata | aaaaatatct   | aaaagccaga | tttcctaaaa  | aaagcggggag | 10620 |
| ctataaaaaa  | aattatcgct  | gtacttgatg   | aatatccctt | gttcgacaag  | ccaattttttg | 10680 |
| atttctatatt | gtggtcttct  | gattactatc   | atcatcctat | tggtgagggt  | tttaatacct  | 10740 |
| ttatacccac  | cgaattaaga  | aaaattaata   | ataaaaaaat | tgaagcttta  | agagaatttt  | 10800 |
| ctgaatatct  | agtaaatgag  | gatgataaga   | aattcgattt | aaccaaggat  | caagaaaaaag | 10860 |
| cagtcaaggc  | gcttttctaaa | tctaaaggat   | tttcacccac | tttattatat  | ggagtttacag | 10920 |
| ggtctggaaa  | aacagaaagt  | tacttaagag   | tgcccgaaac | ttttattaaa  | aataataagt  | 10980 |
| cagtattagt  | tttagttcca  | gaaataaaatt  | taacccccca | attgctctctg | cgattttgaga | 11040 |

|             |            |             |            |            |             |       |
|-------------|------------|-------------|------------|------------|-------------|-------|
| atagatttaa  | tggtgaaatt | ggcatatatc  | attctaagca | aacagcagct | aagagattaa  | 11100 |
| agacttggct  | aaaagctaaa | tttggttcta  | taaaaataat | agtaggaact | cgatcttcgg  | 11160 |
| ctttagtgcc  | ttagataaac | attggtttaa  | taattatcga | tgaagagcat | gaccaatcat  | 11220 |
| ttaggcagtc  | agaagggttt | aaattctctg  | ctagagactt | aagtataaaa | agggcacagc  | 11280 |
| ttgcagatat  | tccaattatt | ttgggacag   | caacccttc  | gctgcaact  | ttaaaacttg  | 11340 |
| taaaagaaaa  | taaatttata | agagttgata  | ttcctaatac | agttgatgga | aacaagcctc  | 11400 |
| ctaaattaat  | agccttagat | atcaataaca  | gccctttaat | aggcggagtt | gctaaaagaga | 11460 |
| caattgaagc  | aatgcaatca | accatagaca  | gaggagaaca | ggttctagtt | tttattaata  | 11520 |
| gacgaggatt  | cgctccactc | tatcaatgta  | gtagttgtgg | gtgggtagca | gattgtaaat  | 11580 |
| cttgtgatac  | aaatttagtc | ttccaccagg  | caagaaatag | attaatttgt | catagggtgtg | 11640 |
| aatctgccta  | ctctgttaat | ttgtcttgtc  | cggcatagca | gtctaataat | gtctaataat  | 11700 |
| atggagctgg  | aacagagaga | gttgaagaag  | ttcttaaaag | cagctttgta | aagactccaa  | 11760 |
| taattagagt  | tgatcatgac | tcaacaaaaa  | aagtgggagc | tatggaggct | atagttaaaa  | 11820 |
| aaattcattc  | ctcagacgca | gcaatttttag | ttggaactca | aatgcttgca | aaaggacatg  | 11880 |
| atttttcctaa | agtcaccta  | agcgttatatt | taaatgctga | taatggcctt | ataagcccag  | 11940 |
| aaattaatgc  | attagagaaa | atatctcaat  | tgcttattca | ggtctctgga | agagcaggaa  | 12000 |
| gaaataataa  | tcttgcaaaa | gttattattc  | aaacaagata | tcctgatgat | ataaatctta  | 12060 |
| ataaaaattaa | gacaggagat | tatatgaaat  | ttgcttctca | atgcctaagt | accaatgagc  | 12120 |

|  |       |
|--|-------|
| aaatgaactt acctccattt actactttat gtctgcttag gtgctcatca ccaactcaaa  | 12180 |
| agagtaatgt agatttccta gagaaagctg ttttaatttt atccaatagg actgatataa  | 12240 |
| atgttattgg tcccttgctt tcattagttt cgaagtcgaa aggaaattat aggcaccaag  | 12300 |
| tctatatcca tgcaccaaaag aagacttttt taaataaggt attaaagttt ttgacaacag | 12360 |
| agtttgaaaa atggccggaa tctaataagg ttaagtggtc ttctgacatt gatccaatag  | 12420 |
| acttaagcta aatattaatc ttaattaatt gtcctgggta tattgggtta ttgtttagtt  | 12480 |
| tatttctctgt attaatttct tctacagtca cccaaatct tctcgctatt tctgataaga  | 12540 |
| catccccctt ttgtattttg taagtcacaa agcctggatc aatactcata aaggtatttg  | 12600 |
| cttttaggtt gtccttaaaa tagttatgta ttctaggaa aacagatctt gcaatcatct   | 12660 |
| ttctccctgg cttgcccttt aatcttttgt cgtcttcagg gttgggtata aaccctgact  | 12720 |
| caaccaatac agaaggata tcaatagact taagaactct gaaatcagcg tactcaacat   | 12780 |
| tcttttttatg aatttttttg aaagggctct ttttaagttg atccaatata ttagttccta | 12840 |
| aaattttact ttcagaaaatt tttttcttat atatttctgg ataggtttg cttgccgcat  | 12900 |
| cctcatcaaa atcaactggc tttagatttt ttatatcagc ctgtattctt tcccttttgtt | 12960 |
| ttttagataa gttccttagca acagtacttg aagcttcacg tgaccatata aaaacagaag | 13020 |
| caccctttac ggaagataat ctaaacccat cagcatggat tgaaaacaaaa atatctgctc | 13080 |
| catatttctt agcatcctgg tatctattat ttaaatctaa tgtctcatct ccattctctaa | 13140 |

tcatcaccgg tctgtaccga taagtatatc ttaaggttct ttctaactcc ttcgcaataa 13200  
 gtaacgtcac atctttttct aaaatatatt ttgggccaac cgcaccaggg tatttaccac 13260  
 cgtgacccgc atctatagca acaacaatat ctctaatact ttatttgagg tttttatttc 13320  
 ttttcacctt aagctctaatt ttatatattt ctgtatttat agtctgagtt gggttttgccc 13380  
 aatggactga ttcatataaa tcaacaacga ttctggtaag acttccatcc tgagatgctc 13440  
 ttactttttt gatgggtag ttgtatggaa catttatctc ggtcggaagg ctcgattgat 13500  
 taatttccat gacaattcta gagggatttt taaagaataa gaccttacca atgaaaacttt 13560  
 atctagacta aagttaatac taatttcatt gttaccctatg tcttgtattt catcgaagaa 13620  
 aacatcattt ccgctgataa aaaaagatag aaaaccaagg attgccaatc tattcataag 13680  
 tttttaacca ttctttaaag ccattgtcac cagaaattaa agaaacctct ctcccttcag 13740  
 gaagatggct gaaaattatt tttagatcaa aacttcttct atgttgaagc ctttcaggcc 13800  
 actcaattag cactacttct ttagagtta tttttctact tagatcaaat atatcaatat 13860  
 cttctgcttc gttagtctta taaagatcaa tatgcaaaaa cattaaaatta ttaaaatcat 13920  
 attcctcgca gagagtgtaa gttggacttt ttacaagatc ctcccatcca caatttttta 13980  
 taatagatct agatataaag gttttgcctg ctctaagtc tccctcaaga tgtattttcaa 14040  
 tttcttgoga agaagattta agtatctcca tagctatttt tgaccctagt tggtttgtag 14100  
 cctcgtcatt tataagagta agtttttttca tcgatttatt aattgtctta ttataggtat 14160  
 taaactagag gcatttaagc ctatttcacc aatatctacc taaaatttta aacctgcctc 14220

|             |             |             |            |             |             |       |
|-------------|-------------|-------------|------------|-------------|-------------|-------|
| cgaatgaaca  | gcaacagcaa  | taatgctcgc  | atttcttata | tcaagacctt  | gggcaagaag  | 14280 |
| agcagtcaaa  | acaccagcca  | gaacgtcacc  | agttcctcct | gaagcaagtt  | caggtccgcc  | 14340 |
| gcacgcgc    | ataaaagact  | tgttatttgt  | tgagtcaaa  | accaccgttc  | ccattccttt  | 14400 |
| taatatgaca  | atagaagccc  | cataagtgtc  | tgcaatttgc | ttggcagcag  | aaatcctatc  | 14460 |
| tctttgaact  | tcttcaatag  | atatgttttag | taatatagcg | gcctctcctg  | gatggggcgt  | 14520 |
| catcaagatt  | gttttatagg  | attttttgat  | taaagatttt | gatgatgcca  | cgatgtgtaa  | 14580 |
| agcacctgca  | tctaataataa | ttttttgaatt | atttgcagac | tttaatatatt | taccaaatat  | 14640 |
| agttttttgca | aaagcagtat  | tagcaattcc  | tggtccatac | aataataacg  | tatgatcttt  | 14700 |
| aattttttaaa | tctatatcga  | ccccttcaac  | catatcaaca | ccaattgcca  | ttacctcagg  | 14760 |
| gttttcttaat | aaagatgggc  | ttacattaga  | cgtgtctgtg | acgagagtta  | ctagtccaga  | 14820 |
| gccacaaaaat | aaacttgcct  | cagatgcaag  | tattccggcg | ccaccatgc   | ctggggagcc  | 14880 |
| agcgcatatc  | aaaacttttc  | caaaattacc  | tttatggcta | tcctttgtct  | tattgggcag  | 14940 |
| aagattttttt | aattcttgaa  | aggtaaaaact | ttgcaatata | gacatgaaat  | tacttaagct  | 15000 |
| aaattatgtt  | gaatgctaaa  | aattcattct  | tcagcgaaat | aaccaatcct  | atcggaatgg  | 15060 |
| aatgatccca  | taaaaaatttt | ttcgtcatga  | ggcactgcaa | tagtgggtag  | gccaaaaaact | 15120 |
| gttctagaaa  | atgaaaaactt | atcttctctc  | ttaagtgttt | tttgattaa   | tttatgaatg  | 15180 |
| gagaagggta  | gagaacaatt  | tgtttttttct | gcacaattac | caaaatcatt  | agcctgtaaa  | 15240 |



|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tctaaagaag  | ttaaccaaac  | cgaaccgtct  | tttaaaaaaga | tattatctgg  | actctgaata  | 15300 |
| aaataactac  | ctgttttagt  | attttgatta  | atatcgtaaa  | cagataaaatt | gtcaccttga  | 15360 |
| ttatagttaa  | cataaagaag  | tccagagccc  | tcactctagca | aaattccatt  | aggcccactt  | 15420 |
| ccatctgttc  | catcaacttt  | atataaatta  | ttgtcactcc  | aaagtactac  | atgccctgaa  | 15480 |
| atacttttaa  | ataatgaatt  | cattaaccat  | tctttcatgc  | tgatatctct  | tttatacata  | 15540 |
| tgagatgcat  | aaaaacttcc  | atctttttta  | agggcaacat  | cattgaaata  | atattggtca  | 15600 |
| gggaccctaa  | tacagccccc  | ccatatcata  | tcccatgaag  | attcattttt  | tataatttca  | 15660 |
| aacatttcaa  | tcgactcaaa  | tggcgaaatga | ttaattacag  | ctagttgata  | gaagcccttca | 15720 |
| tcattttcaa  | caagatctat  | tccgtgggga  | ttgaatatgt  | ctagctcgcc  | tcttatgcat  | 15780 |
| gaagcgtcac  | cccaagagct  | ttctccaaaa  | gtaatttttg  | gtaccctttt  | ttcataggta  | 15840 |
| tttaaatcca  | ttaaggcaaa  | gtaaccaggg  | gtatgttctg  | cataatggacc | gatccctcca  | 15900 |
| aattcagaaa  | taaagaagaa  | tttattatca  | ggcgtaatca  | caatatcttc  | tgggttttgag | 15960 |
| aaattacaaa  | taaccttgat  | acggtcatcg  | gattcacact  | tactaatatc  | catttggggg  | 16020 |
| cctatatagt  | cagtagatac  | gacagttaca  | gaaataaata  | aataaagaat  | agaaaaccggg | 16080 |
| acggtaattct | tataaatagt  | ctttataaaa  | atttctaaaa  | tctttgaagc  | atgattttgga | 16140 |
| agtgcgatca  | gccaaaacccc | ttttaaaaaa  | gatagagccc  | ccataacaac  | aaaaattact  | 16200 |
| gaccatagcc  | tgtcagacca  | ttccggctga  | agaagaccag  | ttaaaaaaaa  | aatcatccca  | 16260 |
| aagaataaag  | ctaaatagcc  | tgagattttg  | actttcgatc  | caacaaaatt  | cgaaaaagaga | 16320 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tatgtataaa  | gaggcttttat | aagaactagt  | aaaccacagg  | caaagaaaa   | gaatgctaag  | 16380 |
| tagtaattca  | taagttagtt  | tttatataaa  | tgctcccttaa | taataactaac | aagtttctaag | 16440 |
| ggcttgtcca  | atggaacatg  | gtgagcagct  | ccaggaaccc  | cctcaaaaagt | cataatgtca  | 16500 |
| ccatatgtat  | ttttaatat   | gtccaagata  | cttccggagg  | ttaataagct  | gtcttcaccg  | 16560 |
| tggatgaaca  | aagcagggca  | gccaaatgaa  | aatgtgtaac  | cgaataaacct | ttcaagactg  | 16620 |
| ctaaacatga  | catcatcaaa  | tttccatctc  | caccagcct   | caatattttt  | tactgagtgc  | 16680 |
| tcagcaatgt  | atcttaagta  | ccaatcattc  | gtacaatctt  | gcttaggcat  | taacctaaac  | 16740 |
| cttttaataa  | tatctgtctt  | gtcttgatag  | tgcttgatca  | ttctgagagg  | agaagagtgt  | 16800 |
| tgattcgggt  | cataatccgg  | tggtcttata  | aatgtatcaa  | taataataat  | attattttatt | 16860 |
| agatccctttc | tttcagatgc  | aacgtaacca  | gcaacatgtc  | cgccgaggga  | gtgtccaaca  | 16920 |
| ataaaaaatat | ttgaaatatt  | ttttttatcc  | ttttcctttt  | caattacaga  | gacaatacat  | 16980 |
| tctccaaaat  | ctttaatgcc  | atatgaatct  | ctaaaagaag  | agtcaccccat | gccaggaaga  | 17040 |
| tctattgcaa  | ctatatattgc | gcagtcctcta | aagtggggcg  | caataggatc  | ccaccatttt  | 17100 |
| ttatgagcac  | ctgttccgtg  | aataagaatt  | attaaaatctt | tgctttcatc  | tttggagttc  | 17160 |
| cagctagaat  | aggatatatc  | cccatgagga  | ttcttgataa  | tctctgagct  | aggcttgtcc  | 17220 |
| tcaatggcat  | ctttgaacca  | ctgtggggca  | tgaataatgt  | cttgatttag  | attgttagtt  | 17280 |
| atttccataa  | acagtattct  | aagctataaa  | aaataaaaaat | atgaataaac  | ttaattttaac | 17340 |

gccagcagca actgtttttag tcctaaaagga ttctcctgat gggatggaag ttttgatggt 17400  
aaaaagatca agtaggcctc ccttcggaga cctttttgtt tccccggcg gcaagattga 17460  
cgaagggtgat ttcaataata agatagaaga tttttgtgag ggcgtgactg ataaagaggc 17520  
ctccataaat ctgggattag atcttgagg tctagcatat tgggttgc atgttagaga 17580  
atgcctttgag gaggttgga ttttacttgc taaaaaaaaag agtggggaag atcttgatct 17640  
agatggagtc gataaacata aatatcaaaa atatagagag atgttgttaa ataataaat 17700  
tgatttatat aaaatctgtt tagaagaaaa tttaattcta atgcctcaac aaatagcccc 17760  
tttctcgc attgataaacc ctgaaataga aactaggaga ttgatacac gtttttttat 17820  
tgccccacctc cccaagc atc agaccggaga acatgatggt agtgagctca tagacagtgt 17880  
ttggatttca caaaagaag cgctcaaaaa atctcgttcg ggtgagatgc ctatgattat 17940  
gcctacaata aaaaatttgg aacaatgtgc acaatttgat tcgggctcta agctttttaga 18000  
aaatcagagg aatctctcaa atgaggatat ccaccaatc ctgccaaagt tttttaaaga 18060  
agatggtgag tggaggggtc tattgcctgg agataaaggg tatgaggatc attaaataat 18120  
atggacttaa ttactaaat aacagctccc aaccctggtg ttttcactgg gggtaggact 18180  
aatacttatt tgattggcaa agatgatata acccttgtcg accctggtcc aaatatatct 18240  
gagcatctag atgaaattat caaagcagg gatgggaaaa taaaaagaat ctttgttact 18300  
catacccata cagatcattc ccagccgca ttgcctttat caaaaactct taatgttcca 18360  
atgtacggaa ggctagtaga tggatgaatc tcatgggagg atgaaacatt tatcccagat 18420

|             |              |             |             |              |             |       |
|-------------|--------------|-------------|-------------|--------------|-------------|-------|
| attatttttaa | atgataaaga   | tattattgag  | acagacgaat  | atacgttaga   | agtaatacat  | 18480 |
| actcctgggc  | acgcatctaa   | tcattttatgc | tttttaataa  | aagatacgaa   | atgcctttcta | 18540 |
| acaggcgatc  | acattatgga   | cgggtcttacg | gttgtttattg | ggccaccaga   | tggcaaatatg | 18600 |
| acaagctata  | tcaattcatt   | agaaaaagta  | ctagattttg  | atattgattg   | ctttgcgcct  | 18660 |
| gggcatggaa  | attatatattca | tgagcctgag  | aaaaccattc  | aatcaattat   | taggcacaga  | 18720 |
| ctaacaagag  | aaagaaaagc   | tcttagaaaag | ctaggagagg  | caggaatctc   | atcattagat  | 18780 |
| aaacttacta  | agcttgttta   | tgatgatgtg  | tcagagatgc  | tccatcctat   | agctaaatat  | 18840 |
| agtctagaag  | cacattttatt  | aaagcttata  | gatgaaaaa   | aagttaaatt   | agataaaagat | 18900 |
| ctattcgaaa  | taatttaatc   | cttttttattt | ttatgtaaga  | ctttctcctc   | aatagcttct  | 18960 |
| atatcaatat  | catcaattga   | gtcttcgtta  | ttatcaggta  | tctttttttac  | atcttttttca | 19020 |
| attttaaggt  | cgataggaga   | ggctccaaga  | tcaaaagtta  | gctcccttac   | atttttttgaa | 19080 |
| atagtatcct  | caacacagtc   | atcttgatcg  | tatgcctcac  | gggtctctcc   | tttatcatcta | 19140 |
| atgggaaagg  | gtctttgggg   | agggccatt   | tgcatgcatt  | taatagtagc   | aactgggtgag | 19200 |
| tagtaatcgt  | cactagaata   | taacttatca  | agttctttttg | gtgatatattga | gcaaccaata  | 19260 |
| attccaaaata | gaataaatgg   | cgctaagagt  | cttttcataa  | actttgtttg   | ttttctataa  | 19320 |
| gagactcaac  | aaccgaagga   | tcggcaaggg  | ttgtcgtgtc  | tcctaaatta   | gatagatcat  | 19380 |
| tctcagcaat  | cttttctaaga  | attctgcgca  | taattttttcc | cgatcttgtt   | ttaggcaagc  | 19440 |

|             |             |             |             |            |             |       |
|-------------|-------------|-------------|-------------|------------|-------------|-------|
| ctggagcatt  | ttgaattaaa  | tctgggttttg | caatagctcc  | aatttcttta | gcaacaaaatt | 19500 |
| gtttcaattc  | ataactaaag  | ttgtcatcaa  | atgattcatt  | tatcattaaa | gtaacaaaaag | 19560 |
| catatatccc  | ttgcccccta  | attggatgat  | caaaaaccaac | aacagctgct | tcagcaaat   | 19620 |
| taggttgagg  | cacaagagca  | ctttcaattt  | cagctgtacc  | tagtctgtgg | ccagaaacat  | 19680 |
| taagaacatc  | atcaactctt  | ccggttatcc  | agaagtatcc  | atcctcgtct | cgcctggctc  | 19740 |
| catcaccagt  | aaagtaaata  | tctttataca  | tacaaaaata  | ggtgtcgatc | attctttggt  | 19800 |
| gatcaccata  | aatacttcta  | attgactag   | gccaaagattg | ctcaataact | aaattacctg  | 19860 |
| cattagagcc  | ttctagcgtg  | tttccatgct  | cgtcatagag  | agatggctta | actccgaaga  | 19920 |
| agggcagagt  | tgctgaccca  | ggttttgttg  | gagtaatacc  | cgctattgga | gagataagta  | 19980 |
| cagaaccagt  | ttcgggtttgc | caccaagtat  | caataacctc  | gcaattagat | ttaccgacaa  | 20040 |
| cactgtagta  | ccaatcccac  | gcctctggat  | taattgggtc  | tccaactgta | cctaaaaattc | 20100 |
| ttaggctatc  | tctttttgtc  | tttttaacag  | gatcatcgcc  | ttgggccatc | agagctctaa  | 20160 |
| tggcagttgg  | agctgtataa  | aaaataactaa | tgtcatgctt  | atcgcatatc | tcccaacacc  | 20220 |
| ttgatgctga  | agggtaggtt  | ggcactcctt  | caaacataag  | tgttgttgct | ccatttgaaa  | 20280 |
| gagggtccgta | caagatataa  | gtatgtcctg  | ttatccatcc  | cacatctgca | gtacaccagt  | 20340 |
| atttgtcctc  | tggtccttatt | ccaaaaagat  | atttgaaact  | aatatgagcg | cctaatagat  | 20400 |
| aacctgcagt  | agtatgtaga  | acaccctttg  | gcttgcctgt  | agagcctgat | gtatagagaa  | 20460 |
| tgaaaaagagg | gtcttcggaa  | tccataggct  | ctggagcaca  | cttattagaa | acatcttttaa | 20520 |

caagatcttc ataccaaaaca tctttttttat catcccaatt aatttcgcca ccagttctttt 20580  
 ttataaccag tgtatttttta acatctggac agcccagaag agcctcatct acattagatt 20640  
 taagtggcac tttttttgccg ccccttaaac cttcatcagc agttataaca attttacaat 20700  
 cagcatcaag aattctatct ttgagtgatt ctggagaaaa gccaccaaaag acaacagagt 20760  
 gcacggcacc tattcgcgtg caagcaagca ttgcgaacgc agtctcaatg atcataggca 20820  
 tataaatata aactctcgag cccttttgaa caccaggtc ttttaaaaca ttagcgaact 20880  
 tacatacttc gtcatggagc tcttttgtagg tcaattcttt agaatcagca gggtcatctc 20940  
 ctcccatat taatgctatt ttgtttggat catttttctaa atgcctatcg atgcagttta 21000  
 agctaatatt tgttttacca ccctcaaacc acttcgcatt attaaattga ttattgaatg 21060  
 ttgttttgaa gtcttccatc cagcttatgt tttcgtttgc tagatttttta aaaaattttag 21120  
 aaggatcttc tatggattgc ttgtaaaagt ctttgtattc atcaaagtct tttatataag 21180  
 gattacttga atgtttttggg ctataaaagct ttcgaggcat tcttaaaataa ttgaaggttg 21240  
 ggggttaatg aaattctttc ctttgggatt ggacattatt tttgtaatga gcgatttcgta 21300  
 atcgctatca ttggttttcta aattaatatt tgcagcatta attattaata aaggggcaga 21360  
 actatagtct aagaaaaaac ttgagtatgc atcattcagt ctttccagggt agtcaagagt 21420  
 tagatatgtg tcgttaatat ttcctctctt agtaatcctg tcttttaaca catcaatagg 21480  
 tgcctgaaga tagattacta ggtcgggtgt tggcgcgtcc agggttagat ggtcatatac 21540

|           |          |          |        |         |        |       |       |        |            |       |
|-----------|----------|----------|--------|---------|--------|-------|-------|--------|------------|-------|
| tttgtcata | agatccat | tt cctcg | tttaga | aagag   | taacc  | tcagc | gaata | atcgat | cttt       | 21600 |
| ttctattaa | aaatcag  | caa cctc | actgt  | ttcaaaa | aagg   | cttct | tttgt | taagat | cttg       | 21660 |
| aatttgttg | cat      | tttgaa   | acaaga | agaa    | aagct  | gagt  | g     | ctag   | agctg      | 21720 |
| gtttttata | aaattct  | ttta     | agaagg | att     | ctctg  | ccggt | tg    | ttcta  | aaaa       | 21780 |
| attaaa    | tg       | tgctat   | ct     | tattg   | ctaa   | agtag | ttttc | ccaac  | accta      | 21840 |
| aattgcg   | ata      | tattttg  | gaa    | gtggt   | acttc  | ttta  | atag  | ct     | gggttcatt  | 21900 |
| gtttttat  | cc       | tagatt   | atct   | caact   | cg     | tta   | aga   | gccaa  | atatttt    | 21960 |
| tcacttt   | ttt      | cattcat  | ctc    | agca    | ataat  | g     | ttcct | ctca   | atgag      | 22020 |
| aaatcag   | tc       | accact   | ttccc  | aata    | ccactc | ttgt  | ctcc  | ag     | aagctctctc | 22080 |
| agaaaa    | atcat    | atgcag   | ctct   | aaac    | cttgga | tg    | tcga  | agag   | ttttat     | 22140 |
| attctg    | ctat     | gaag     | ttttta | ttga    | agtacc | caa   | ata   | tcct   | taat       | 22200 |
| cttggtat  | tg       | ctgat    | ttt    | ttgt    | gttcg  | cgaa  | gtac  | ac     | catcc      | 22260 |
| tttctcac  | at       | taatt    | ctctc  | attct   | tagaa  | cact  | ttttc | a      | ataac      | 22320 |
| gcagccata | aa       | gaa      | gcctgg | tg      | taatt  | gac   | tg    | ttgg   | ttt        | 22380 |
| cttagag   | cat      | gtgc     | ataac  | attg    | ctgg   | ca    | aag   | tcatt  | cc         | 22440 |
| agata     | tttat    | ttaa     | atg    | aa      | tgaa   | gag   | att   | ttc    | aaat       | 22500 |
| aaaaat    | at       | ttt      | gcaga  | actc    | atca   | aa    | caat  | ctag   | cg         | 22560 |
| cc        | tttat    | cat      | agat   | gg      | catc   | ctta  | aacc  | ga     | ttat       | 22620 |

|             |            |             |            |             |             |       |
|-------------|------------|-------------|------------|-------------|-------------|-------|
| aatcttatag  | ctcttaagct | tctcactgga  | tcttcttcga | atcgtctttg  | gggatctcca  | 22680 |
| atagatacaa  | taaccttctt | gtgtatatgc  | taaagcccat | cgttatgac   | ttctattttt  | 22740 |
| tttgtaacag  | gacagtagta | aagggcattc  | acagtaaaat | ctcttctatg  | acaatcttgt  | 22800 |
| tcaagagtgc  | cccaattatt | atctctaaga  | atctttcctg | tcgaatcagt  | aacaatatatt | 22860 |
| tcaccatctt  | cctgatcact | tcctgatctg  | aaagttgcta | cttcaagcaa  | ttcactccta  | 22920 |
| ttaaaaacat  | gaaccaattt | aaatcttttg  | ccaattattc | ttgaagcttt  | aaatgttttc  | 22980 |
| ctgatctgct  | caggcgtggc | attagtggtt  | atatcgaaat | cttttggctc  | tagtcccgtt  | 23040 |
| agtgcatcgc  | gaacgcaccc | accaaccaga  | taggcctgaa | aattattttt  | ttgaagggtct | 23100 |
| tgaacgacag  | atatagcaaa | tttacttatt  | ctattattat | ctatcaaatc  | taattatgaa  | 23160 |
| tttttatgat  | ttaagggcat | tagccattta  | gctgcttttc | tttaattttcg | tcaagtgttt  | 23220 |
| tgcaatgaat  | gcaatgagtt | gcagttggtc  | tagcttcaag | cctcttaatt  | ccaatttcat  | 23280 |
| caccacagga  | ttcgcaccaa | ccataatcat  | cttgcttaat | ttgttcaata  | gatagaccta  | 23340 |
| ttttgctgat  | aagttttctc | tctctatctc  | tggttcttaa | ttcaaaaagca | aactcttcct  | 23400 |
| cttgagaagc  | cctgtctact | gggtctgcat  | aggtttcacc | tttagctcta  | agatgatcaa  | 23460 |
| aagttttttg  | catttcatcc | ttaagatgtt  | ctttccaaag | aagaagaacg  | gcaacaaaaat | 23520 |
| gtttcttcat  | tgctgcactc | atatattttt  | cacccttctt | agatttataa  | ggtgcaattt  | 23580 |
| tagattttatt | attagctatt | gttgctttttg | cagacttctt | ggctgcaact  | tttttaactg  | 23640 |



|            |             |             |            |            |             |       |
|------------|-------------|-------------|------------|------------|-------------|-------|
| gagctttttt | ggtaaacagtt | tttttaactg  | tagctttttt | ggcaacagtt | tttttaactg  | 23700 |
| gagctttttt | tactgtcttg  | gatttttttt  | cgaccatgta | agattttata | gaatttttagg | 23760 |
| gtggagaaaa | tatcagatac  | tgacaaaaatt | agctagtcac | tttttaattt | atttaatact  | 23820 |
| ttcaagtacc | catccgagct  | aagtcttggc  | ccaaatgtct | caacaaacct | ggaagatgca  | 23880 |
| tagtttgcaa | acttagcaca  | tgcttcaata  | ttattccctt | gaaggtaggc | atgcataaac  | 23940 |
| gatccggcaa | acatatcacc  | agctccattg  | gtatctattg | gagttatttc | ttctgcttga  | 24000 |
| gcatgcttct | caaccctttt  | atctataaca  | acacttccat | cggcgccctt | tgtaatatgca | 24060 |
| gtcatatagg | gctttttctt  | ataaaagcta  | acagcatcat | caaggctttc | tttaccagaa  | 24120 |
| aaagcaacag | cttcatcatc  | attacagaag  | atcatatcta | ttccatatga | ctctatttaa  | 24180 |
| tcaaatTTTT | ctttaaaacc  | atgaacaata  | cctgcatcag | aaagagacaa | ggcttttcttt | 24240 |
| acgtccttgt | ctttaagggt  | ctctaagact  | gaaataacag | cattaaagtt | atcgtcactt  | 24300 |
| gttaccatgt | agccttcgat  | ataaaaaatt  | tttgaatttt | ctacaacatc | aaaatctata  | 24360 |
| tctgatttac | caagatacgc  | actaaactcc  | agcatgctgc | tcatagttct | cttagcgtca  | 24420 |
| ggagtaacta | aaattaagca  | tttcccagtt  | ggttgatctg | tattttcaga | gctgacacca  | 24480 |
| atatgtttga | ctccagccga  | cctgagacta  | tcaagatagt | ttcttccatc | ttcatcatca  | 24540 |
| gaaactctgc | atacatgatg  | gcaattcgaa  | ccataatttg | ctgcagcaac | aagagaattg  | 24600 |
| gttgcagagc | caccgcaatc  | agaaatcgat  | tcggctccca | tttcaataag | tttgctaatt  | 24660 |
| ataggtgcct | gttcttcaga  | agatgaaaga  | gtcatagagt | cggctacaag | gcctacactt  | 24720 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| gataaaaatt  | catggctgac  | tttatattga  | gtatctacta  | aagcatttcc  | aagggcgcta  | 24780 |
| atatcatatt  | tcatgtgtta  | ttccttttttg | tattatttgt  | tttactgttt  | ctaacgtttct | 24840 |
| attaatttct  | tgatccttat  | gcattgctga  | aatgaagcct  | gcttcataatt | tagagggggc  | 24900 |
| aaagtataatt | ccacttctaa  | tacatgaatt  | taaaaaattt  | gaaaatagta  | catcatcagt  | 24960 |
| ttttgcaaca  | tcattaatat  | tattagggag  | ttcttctgaa  | aaaaagaatc  | cgaacattcc  | 25020 |
| accaattctg  | tttatagaaa  | acggaatacc  | tgactcaatc  | attaaggttt  | tcatcccatc  | 25080 |
| caaaagcaca  | gatgcatttc  | tttctaactc  | tttaaacgga  | ttttcttttaa | tcaacaattg  | 25140 |
| caacaaaagca | gttcctccag  | ccatagctag  | tggaattgcca | gacaaagttc  | ctgcttgata  | 25200 |
| gacaggacca  | gaaggagcta  | gatagttcat  | aatttcttct  | ttgcctccaa  | aagctccaac  | 25260 |
| aggaagaccg  | ccacctatta  | ctttccccaa  | agcagtttaag | tcaggagtaa  | tattataaat  | 25320 |
| ttcttgagct  | ccgcctagcg  | aaactctaaa  | gccgctcatg  | acctcatcaa  | atattaaaat  | 25380 |
| agaattattg  | gctgagggtg  | tttccctcaa  | taacttttaag | aaatcttcat  | gacctggaac  | 25440 |
| aaagcccata  | ttccctgcta  | ctggttcgac  | tattacggcc  | gctaagtcat  | cttttatctc  | 25500 |
| attaaaatatt | tctaaaaaact | gttcttttatt | attgtatttcg | caactaaatg  | tatatttttgc | 25560 |
| caaatctgca  | ggaacccctg  | gagagtcagg  | taagccaaaag | gtggcaacc   | cagaacccgc  | 25620 |
| cttaattaaa  | agagagtcaa  | catgaccgtg  | ataacagcca  | tcaaattttaa | taatttttatt | 25680 |
| ccttcctgta  | aaacctctgg  | ccaatctaata | cgttgtcatg  | gttgcttcag  | tacctgaatt  | 25740 |

aaccattctt atttttttcaa ttgaaggaat gcattttttta attagcctgg ccacatcaga 25800  
ttcaagactt gtcggggcgc cataactagt tccaagcgca acttgatttt taattgcacc 25860  
tacaatgtct ggggtgtgaat gaccattat catcggaccc caagacccaa tataatcaat 25920  
atattcatta tgatcagcat cataaaggta tgcgccggat gctctttcaa aaaatatagg 25980  
attgccattg atatttttga atgcccttac tggngaattt acccctccag gcataagagt 26040  
cttggcctct ttgaataagg ctatggattt atcaattttta ttatgtgtca aagctaattc 26100  
ttcttaattt ttaatttcga tatgatatca acttaattct tattatgtat ttaattgttt 26160  
tgaaaaattca tccacagtat tccagttggt gaactcataa gtatttgtaa catctgtagg 26220  
accctttgta atccacataa tcaatctaata catgtgttta tctataaatt tatatttttg 26280  
ataatctatt ttgcctgcaa agactgctaa tttcattggg ttccaaggag acagctctaa 26340  
aaatttttgc atatacgggt ttgttttcagg tgtatttttt tcaggctttc tcgcaactac 26400  
attaactgaa aagaaggcat tttcttttgt ttcaagacaa gcaacatttt tttgaataaa 26460  
ttcataaaagt tctggtttgt gtttgccata cctaattgtc gcaccaataa taatttttctc 26520  
aaattgatat aaatctaact ctactgcttt tgcaatatgt attatttttg aagattcaga 26580  
cacatctaata ttagagaaaa ttttttttaca aatttccaga gtttgcccat cggttgtaga 26640  
gtagattagt agagttgatt tcataaaatat attttatatt gttaattaaa aaatttcatg 26700  
gcacgcgcaa aatataaaca aaaatatagt atatttaaaa aatgattaaa tattttattta 26760  
ttatcgcctt attcatcagc aactttgcat atgcaaatat tgatgcagct aggtgcgcag 26820

|   |       |
|---|-------|
| ggattagttc agatcaagaa agacttgatt gctatgattt aatttttaaa gctaatgatg   | 26880 |
| agctaccttt agatagtaat attaaactc tcatcacacc agctatcaag gctgtgaccc    | 26940 |
| cagctgattc aataaaaaatt gaaaataaaa caacaaaaa aaaggatttc gggcttccaa   | 27000 |
| aaacaaaaat caaaaactct gcaaagaact caataaaaac ctcggtagta aggattaaaa   | 27060 |
| aaacaaaaag tggtaaatata atttttactt tagaaaaatga gcaagaatgg actgctgaaa | 27120 |
| cttcttatag agcaagggaat atgtttaaac cagaaaaccgc agtcatttta gaagaggccc | 27180 |
| tggttagtagg ttttttatatg attaatataa gtaataaaca gaaaattaga ataaagaggt | 27240 |
| tgaaatagca atgaccatac agagtatagc aacgaccgag ggtgccatga aaagaataag   | 27300 |
| atctgttctg gggccccaaag atggctcaag cttcagggtt tatgttaccg gaggtgggtg  | 27360 |
| ctctggggtt caatatggct tcaagtttga taacgatatt gcattcgatg atgatgttat   | 27420 |
| aaattgtgga gatttctcgc ttttaataga ctctatgtcg taccctatc tttatggatc    | 27480 |
| aactctggat tttgttgaag atctctcagg ggctaaattc gttattaaaa acccaaatgc   | 27540 |
| caaaaacaaca tgcgggtgtg gagagtcatt tacagtttag attttgtaat tgaacccaaa  | 27600 |
| agtcctttta cattcttgct tgatgtctga actatcagag cctcattact gactctttta   | 27660 |
| tatcccatcc aggcaaatgc catggactca atagcaaaaa catcatggcc taaatcactt   | 27720 |
| gagaggacga tatcattact agccatctca gaaattcttt ttaccagata cttgttatgg   | 27780 |
| gcaccaccgc cacaaatgac aatgtcacaa ttattatgac catttttatg gattgaattt   | 27840 |

|             |             |             |             |            |             |       |
|-------------|-------------|-------------|-------------|------------|-------------|-------|
| attatagatt  | ttgctgaaaa  | ttcaaccagg  | gtacatagaa  | tatcttcagc | tttctttttt  | 27900 |
| aagaatttct  | tggataatat  | ttttatatta  | aagagctctt  | tgcctgttga | ttttggacat  | 27960 |
| tttcttttaa  | aaaaattatt  | ttgaagcaat  | cttcctaact  | caatgtgac  | tactttttcct | 28020 |
| ttggctgcaa  | ttgcaccatt  | tttatcaaaa  | ggaatctgaa  | gaaaatcact | acaatatgca  | 28080 |
| tctaaaaattg | cattaccagg  | gcctacatca  | gttccccata  | tgtcatttct | atttttttaca | 28140 |
| aatgagtaat  | ttgaaatccc  | ccctatatatt | aaaataattc  | gtgggtttct | agctttataa  | 28200 |
| aataattgat  | tatggaattc  | cggaacaaga  | gggggcctt   | cgccaccag  | agcaatatgc  | 28260 |
| atattttctaa | aatcacttac  | caccagaagg  | ccggtttctt  | ttgcaacaat | attttggatca | 28320 |
| ccaatttgca  | tagaaaaatgg | aaatctctta  | ttaatttcat  | gcctaattgt | ttgtccagaa  | 28380 |
| atagcaacac  | actctataga  | tgattttctta | attttactaa  | atccaatgc  | ctcattaata  | 28440 |
| gattttgaaa  | ataagaatcc  | tatctcttta  | ttgatagtgc  | ctaaatctga | caacgagctt  | 28500 |
| tcgttathtt  | caataagttt  | ttttactttt  | aatcttaggg  | attttgggaa | tttaattgaa  | 28560 |
| tggaataact  | caagatgtat  | ttttgtacca  | atgctaagaa  | atgaaatatc | aatagcgtca  | 28620 |
| tgactagttc  | ctgtcatatgc | tccaatatag  | atctttttac  | tcattactac | ttggagaaaag | 28680 |
| tttattgaac  | tcctccattg  | ataattttatt | attcttttaat | agtgtatcaa | aacttggcct  | 28740 |
| taaatTTTTT  | ggtactgggt  | cagctgatgg  | taattttact  | tttataggat | ctattctttt  | 28800 |
| cttacctatc  | ttaaaactcat | aatgtaaaatg | agggcctggt  | gccagtcctg | aactcccaac  | 28860 |
| gaaaccaata  | gtatgcacct  | gagaaacttt  | tttgcctttt  | cttataacct | tactaaaattt | 28920 |

ttctaaatga caatatctag tcgaatactc atttgtatga ttataaacta tctcattacc 28980  
 acatccattc ctttgaccag aaaaagaaac aatcccatcg cctgtagttc tcaccggcga 29040  
 tcctcttttt gctgcataat cgactccatt gtgagctctt attgtatgaa gaacaggatg 29100  
 cattctgttt ggattaaat gagaactaat atatgcaaaa tctaaagggg ctcttaaaaa 29160  
 ggccttttgc atattatttc catttttcatc aaagtactgt tttttactgg cttcagtgaa 29220  
 aaacctattt gcaaaatagg tattaccatt gttaataaac ttagcaatta cgatatcgcc 29280  
 atttttttact ttctctccat cactataagg agtgtcataa attacatgaa attcgtcacc 29340  
 ctctctaata tcaaaaaaga agtctacatc ccagccaaag atataagcaa agtccataat 29400  
 cacactttcg ggaatatattg catctagagc tgcctcataa aatgaggatt gaataatccc 29460  
 gctattgtaa gattcaatta actcaatact tttgcttata ttcttaata ctattttctgg 29520  
 agtaaggcta attgaaatag aatttatttg atctttcatg atctcaattc tggtagctc 29580  
 ctcaccagaa tattcaaaga gcatttttttc ccctggttta atgtagcaa tgatattttt 29640  
 ggagtcta ctaaaaattt tataagcagt atttaaaggc actgaaaagt tttcaaaaaat 29700  
 tattgaaaga ttttctccat ctgacacctc atgcatctga tatgttttcg tttgctcgag 29760  
 aggtatagag aattttttcag taatctcaat ttcttcaact ggcaaagatt catatgtttc 29820  
 tatatcaata taaagcatta atattaagac aatagaaaaa gcaaaagaaa caagtaccgc 29880  
 tcttttttgg actttttttaa aacctatcat tagtcctctt gaatatattaa gagttcagtg 29940



atatggtctt ctgtttggag gccaacagct tgttttgcac agtatctgag aaaatctata 31080  
 gcttctcttaa tttcatcaat agtatttttgt actgttttttc cggtttcatt cataaggtaa 31140  
 taaattagct cagaaggatt tgcttcaatg tcgtctgcaa ttttttctaa aatggatgct 31200  
 cttttttcaa catgcatcaa agaccattca ctaatatattt gttcttcaag gctgcccttt 31260  
 atataaatcta gatcatcata tgaggctgta ccaatatattc ttccatctgc aagagatgag 31320  
 atgtcatgag tattggtttt cttataatcc ttgcttttat aaatcgatga agcctttatt 31380  
 tcttttgaat caaacctttcc aagctcctct tcaagcattt ctaaattcac cctttcacctt 31440  
 aaatcaaatc ctttagagtt cggctctgtca ttaaatatat ttcttggcat tggatatctct 31500  
 tttttttcat cctcaatttt taaatggggc cctctggcaa gccaaagcaga atctgttttca 31560  
 ggatcaagta acctattaat aaatgagctg ttgctccat tttctaatag tcttctttaca 31620  
 agatagggtg gcaagtcttt atatttgcca atcggcgcac aaattgaggt attttttttca 31680  
 gtattttaaga ttttgtttgc ggacttataa agcagctctc ccattccaaa tagtcttttga 31740  
 aattcataat ccttatttga gccaaagatga tggattgcag aaatggtatg tgcattatgt 31800  
 gtagcaaat tagggtaaat tttttccaca ttaaaaattc tttttgcaca ggctaaataa 31860  
 gctaagtctg taacagattt ttttgtatag acaggataac catcatagcc atagatttga 31920  
 gcatgcttaa tttcataatc ccaatatgct ctttttacga gccgaacatg catagggtgct 31980  
 ctgttttcta gtactcttc taaccaatct attgtggcta tagctctttt gccatatgcc 32040



tgaacagcaa taccaaatcc tttccagttt ttaatatattg gtgaaagggc catctcttta 32100  
 ataatctcta aactaacggc cagtctgtct tgctcctcag catcgattgt aatttctaca 32160  
 tctttagact ttgcatactc tgtgagctga ataaagtttg gaagtagatc agacttttatg 32220  
 tcttttagct ttttcatctc atatctagga gataaagctg atatttttat tgagacacca 32280  
 ttaattgtat tttttgtcag atttatcttg cegacttcat ctatagcatt cttataagac 32340  
 tgataataag tgtcgggcttg ttcagcgttt cttgctgcct ctccaagcat gtcaaatgaa 32400  
 taaatttcat tttctatatt cttaattttt tttatgtcat caaaatctct gcccataaca 32460  
 aattcttgac tgagaatatg catggcacca actacagcat ttctaattgg aaactcgcca 32520  
 gattttgaga ttaaaagaact taataggcg cttgggtttt tggtcactc atcggtgta 32580  
 gaaactacct tgcctgcaag aagaagacc catgttgatg cattaacgaa aacactatcc 32640  
 gctttgttta aatgctctat ccacgcacc tcagataact tctcagatat tattagatcc 32700  
 ctbtgttttt tgtctgggat tcttagtatt gattcggcta aacacatcag agcaaacacct 32760  
 tctttattat ctagcccata ttcgcttaga aatgcatcta gttttgttct ctcgctctta 32820  
 ttctctctgc aagcatcaat tattttattg agcatttttt gaaatcgaag ggtcatttaa 32880  
 aaaatcagaa ttactaataa gatctgaaac tatttcttgt tcaggataaa acttgttact 32940  
 tgттаatgtc atagctatta ttttaatctt atagtgccaa gtatcaactt actgattcta 33000  
 ttaatgactt aaaaacatta tgatagtcac atgagtgatg cacttaaatt aattaaacga 33060  
 ggaaccgacg agatcctcac agagtctgat ttaaaaaaga aattagatc tggaaaacag 33120

ctaatcatTA aggCGggggtt tGatccaaCA gctccagacc ttcatttggg tcataccgta 33180  
 ttattaaata agctaaGaca ttttcaagat ctTgggcata aagtaatttt tctaattggT 33240  
 gatTTtactg gtcaaaattgg agatccctcc ggtaaaaaata aaactaggcc aacacttact 33300  
 tcagaagaat taatttcaaa tgcaaaaaaCA tatgagaaac aagtTTttaa aattctttaa 33360  
 aaagaattaa cagaagttaa atttaattct gagtggtgca acaagcttgg tgcagatggT 33420  
 ttgattggTc ttgcatcaaa atataaatgtt gcaagaatgc tggagcgtga tgattttaaT 33480  
 aaacgttata gcgcaaatca aagcatagct attcatgaat ttttataccc ccttgttcaa 33540  
 ggatatgact ctgtagccct agaggctgat gtcgaatgcg gaggaacaga tcaaaaaattt 33600  
 aatttgctag tagggagaga gctacaaaga tcctatggTc aagaacctca ggttgTTtta 33660  
 actgtaccca ttctagaagg ctTggacgga ataaataaaa tgtccaaatc attaaataac 33720  
 tttatagcaa tagatgaaga gcctaattgat atgttcggta aaataatgtc tatttcagac 33780  
 gagttaatgt ggagatggTt tgagttactc agctttacct cagagaaaga aataggagTt 33840  
 cttaagaaga aaatggaaga agggaccacac ccaagagata ttaagTttct tttagcagaa 33900  
 gagttagtag atagattcca ttcagagggT gatggTtcga aatgcaagga agctTTttctt 33960  
 caaagatttc aaaaaggTca aatgcctgat gacattcctt ccatgtcagt tgatgttggg 34020  
 gctgagggca ttccattagt aaacctttta aagaattgtg agatgacatc aagtacatct 34080  
 gaggcgatga gacttgTtaa acaaggcgga gttaaaatcg actctgtaaa aatagaagat 34140

cctaaaaatgc taattttcaaa aggccaggag tctattttatc aggttggcaa aagaaaaattt 34200  
ttaaaaaatta aaacataatg aaaaaataaat tagtccaaat atttcttctt ctttttatag 34260  
ttgcgtgtaa tcaggatagc cctaataatta aacaaattag taatatgcag tatttttattg 34320  
ataatgaaat aagggaagga atctattctg ttgaaccagg ctgcaatac tcaattatcc 34380  
aaaatggaga tcaaagtctt gaatcaccat tgctgcagga tacaattaca gctcattttc 34440  
acgggaccct cactgatggc tcagttttttt ggagttctgt tgaaatgggt gagcccttaa 34500  
cagtcgaact atcaggtcta atagttgggt gccaaaaaat aatctctatg atgaagaaag 34560  
gtgatgaatg gagagtttat atcgacccaa gtatggccta tggcgatgag ggccggcctg 34620  
ggataccttc aaactcaatt cttatttttg atattgagtt attagatac caaaaaaact 34680  
aacccttata tataacggat agagcgtatc cataaacctc agcaacctgg ttaataattt 34740  
tatttttagg tgttcgggt ccatgaccag ctcttccttc aattctaatt aatatggat 34800  
tatcacaccc ctgggactct tgaagttttg ctgcaaaactt gaatgaatga gagggtaacta 34860  
ctctatcatc tcttttagct gtagtaatca gagtagtttg atagcactca ccttcaacaa 34920  
tattatgcag aggcgaataa gctaataagt ttcaaaactc atctttctta tctggagagc 34980  
catagtcact tccccaggcc cagcctatag taaatttatg aaatctaagc atgtccaaga 35040  
cgccaaactg tggaattgcg actttaataa aattaggatt ttgcaacatg gtcgcagcca 35100  
ctagcaaacc tccattagag ccaccttgaa tagcggttga tgatggagaa ccaattttctt 35160  
gtgcgtgtaa aaatttagca gaatatgcaa agtcatcaaa aacattttgt ttattaaaa 35220

gcctgcctgc atcatgccag ttatcaccat attcaccacc acctcttaga ttaactacag 35280  
caactattcc ccccttggtc atccaagtga ggtagctttt actaaagcca ggtagtcttg 35340  
agatatataa gccaccatat ccataataaa gtattggagt attactgtct atttttaagg 35400  
actttttgta actaagatgg atgggtattt gtgttccatc ttgggatgga aagaatttaa 35460  
agtcagatgt aaataaaagt gagtcgtgtc ctttaagatc ttctttccaa aaaagttctt 35520  
gtgacatatac agttagatta attttgtata tttctcgagg agttacaaaa ttgtataaatg 35580  
aaaaataaga tacctcatct tcaatttcac caccaaaacc acccattgtt ccttttcttt 35640  
ctgttgctag cttattttta tatgtctcct ttaaatcaaa gaagtgaacc tcagtaaaaag 35700  
tatcaacgag ataagaaact acaatagaat tatttataaa gctaacacta ctaatagaat 35760  
tagtactttc accaacgact tcattccaaa caaatgagcc attttttatt gtaagtata 35820  
ctacttttcc gttcgcagca ttttcagttg agtaaaaacca aaaagtatca ttcttgcttt 35880  
ctaaaaagct ataagcgct attagctcat ctattaaagg aataaaaagg tgatcagggc 35940  
ttagctgaac atagagtcta tttctttcat ctgtgccttc accaatagat agaaaacttaa 36000  
ttttagaatc ttttacaaca cttattcccc aactccatct tggcttttca ggattctcat 36060  
aaacaatcac atcctcatct tgagcagtac caattttatg aaacattaac ttggagctg 36120  
tattaatatc ttttaaaagc tcttccgatg gctcgtcgtata tttttggtaa taaaatccag 36180  
aatcatcatt ttcccatgaa gcaccagaaa atttagccca ttcaatccta tcatcaagag 36240

|             |             |             |             |             |              |       |
|-------------|-------------|-------------|-------------|-------------|--------------|-------|
| tttttcctga  | ttcaatatca  | agtactttcc  | aggtttctcca | atcagaccca  | ccatctgata   | 36300 |
| ttgagaaaagc | tagaagagat  | gcatcattac  | taacgctggt  | acttgcaaga  | gagatagttc   | 36360 |
| catctttctga | gaactgattt  | ggatcaagta  | aaactctatc  | ttggcattct  | tcacagtcct   | 36420 |
| taatcatcag  | cttgctttgt  | tgccatgacc  | catcattgaa  | ataataaaaa  | gttttttttgt  | 36480 |
| taacctgata  | aggcatgcctt | atcgaatctg  | tatcccaaac  | ctcatctaaa  | ttcttagcaa   | 36540 |
| tagatttttt  | atatttattt  | tggcctataa  | attttttgtt  | gaaattattt  | tgtctctcga   | 36600 |
| cccagtctgt  | tgagtcctca  | cttgtaaaat  | cctccatcca  | tctataagca  | tcttcgatta   | 36660 |
| ggtatcccatg | gactttcttca | ctaaaaaggaa | ctttatttga  | ttcaggggtat | tcaaaaattat  | 36720 |
| ctgtttgatt  | tgagcaacta  | actagaatta  | atagagctgc  | catgtatatg  | tatttattgc   | 36780 |
| gcataaaaag  | cataataactt | tacaattacc  | ttatgaatca  | attcaatcga  | gcatggcaac   | 36840 |
| ttttacgcaa  | aaaattggaa  | gtttttttatt | gtacttgccg  | caccatcat   | gatgcttgaa   | 36900 |
| atagcaatgg  | catcaatggt  | aaccccaata  | caaaacgtta  | ctcagcctga  | ggatatattcta | 36960 |
| gaattttttca | atgagaatat  | agccttttcta | ggctctgtta  | gtctcttagg  | cgtagtttta   | 37020 |
| agtatggctt  | ttatgggagc  | gcttttttgtt | tcgtatgcat  | ccatagaatc  | agaaaaatgag  | 37080 |
| attgagcctc  | taaatgcctt  | attttttaggc | ataagaaaaat | tctttccact  | tctggggggct  | 37140 |
| tatcttatag  | cttcagttgg  | tgtcttttttt | ggtatcttat  | tattaatact  | cccagctttt   | 37200 |
| tatgtagcag  | caaggctttg  | cattttttcct | gcattttatta | tgcttgaaga  | caaaggagct   | 37260 |
| atagaatctc  | ttaaattatc  | ttgggaaaaag | acggatgagc  | atggcaccac  | tttgtttggg   | 37320 |

ctaccatta ccttcttttc tttaacaatg atttttgcac cagtgccca atccattata 37380  
 agtccgggat taatgcaatt agttgttctt gcaattattg aatatgtaat agtaattcca 37440  
 tggggctatg tataatttttag ttatacaag tcattaaaaa gatattaggc agattaacta 37500  
 gaataaatta atacgaagct ttctgtttac aaaaaaacac accactctta taatccttcc 37560  
 tccgaagatt tgcactttgt taattcatgc aaacaacttt gggctctgtag ctcagcttgg 37620  
 ttagagcgca cccctgataa gggtgaggtc ggtggttcga gtccaccag acccaccatt 37680  
 ctttgaacat taacttttat ttcctttata cttattctta actaaataag gatagtatta 37740  
 tggctaaagg tttagacaag caaaaaaacg acaagaagaa aggcaaaaact ctaaaaagaaa 37800  
 aaagagcggc aaaaaaagaa aaaaagaaat agtttttgt agaagactat tattatctag 37860  
 tcgacctagg gccatcttta aaatctattt taaagaacaa cattattccg gtcactgagc 37920  
 ttactaaagc taaaatagaa aatattttta aaagaagggt gtcgatatta tctctttctt 37980  
 tccaatccat aatatgaaac cccacatca gatcccatat tttccattta ttagatctaa 38040  
 tggccacaac ctctgctgag taaacattaa tgtaaacatt taattcagta cctttaacat 38100  
 ttttactttt tactctataa attggcaatg atctgcctct atattcagat cctgattttt 38160  
 catttacaac ttcttcggta gcaaatggca aaagagttgt ttggcttgaa actgaatcta 38220  
 tagcatcttc catagaaatc tttgttaaa gctgccaag catattcaga tacttagttg 38280  
 agccctttgt tgtgatgata acaatctctt gacctaatct ttttttgaat ttaacttctt 38340

|             |            |             |             |            |            |       |
|-------------|------------|-------------|-------------|------------|------------|-------|
| tagctttctc  | tatttcaaaa | tttaacttac  | ttaaatcaaa  | agatgtttca | acatgattaa | 38400 |
| ggtattgctc  | gcctcttacg | agctcaatct  | tattaaagc   | aaagtatatg | cctgagatag | 38460 |
| tccaaaagtaa | taactgaaga | gagataaaaa  | aacttaggta  | cttatgaatt | ttcctaacta | 38520 |
| aaaagttcat  | cttttcataa | tcttcattat  | ttcatcaatt  | ttacatcaa  | attcgtcatc | 38580 |
| accttttaaa  | gattccttaa | tgcattcacg  | caaatgggtt  | tttagaattt | tgccctcaac | 38640 |
| agtaataatt  | gaattcttaa | gtgcttttat  | ttgattaa    | atatcaacgc | agtattgacc | 38700 |
| ttcctcaacc  | attccttgaa | cgcctcgcac  | ctggccctct  | attcttttta | agcttattaa | 38760 |
| ttgctcttta  | tgacatggat | gtttcathtt  | ataaaaacct  | cggaactata | tacgtaaaaa | 38820 |
| ttgtagcgaa  | taaaaaaatt | gctacagatg  | catagtacag  | atacctagat | ttttttctag | 38880 |
| tttctgaaca  | tgctctgcct | agttctgggt  | cagcagggca  | aggcatatga | taagttttgt | 38940 |
| agttgatata  | tcctgcaatg | acaatcatta  | ctaatagcaa  | tagtgttatg | taaagcttat | 39000 |
| attgagatag  | ggtaattaag | aatggaata   | cagttaccag  | gcttgcaaaa | cttgccccag | 39060 |
| caccaagagc  | aacaaaaata | gctggcaagg  | cacaacatat  | caaagtagaa | gaagatgcaa | 39120 |
| ataatgaaaa  | aaaattcgac | gccttatcat  | tcatagttaa  | taaatttcta | ggatttgaag | 39180 |
| atcagacata  | ttttggccat | tagaaaagaat | cttcttttcg  | atctcagaga | aggtgatttc | 39240 |
| ttttgcctta  | gaaaaggcaa | tcactacttt  | tccattatca  | agatcaacat | ctattctttt | 39300 |
| gacatcttta  | tcttttagaa | aagtttttga  | aataacctctt | gcacagaaat | cacaaaacct | 39360 |
| ttccttcacg  | cttacgatcg | ctatatattac | atttgagagt  | ccttcaacaa | atgcatcaaa | 39420 |

cctctcaggg tttaacatcaa gatctttttcc atcaaccatg gtttcatgta aatgaccttc 39480  
atgagaatgc ccatgcatac cactatggtc attctccgcc gcaataaggc cgcttgaaag 39540  
caaagtaatt aataataaatt ttttcatatt tttctccttt taaaatctgt acataaaaatg 39600  
taaatcccag ctcgatatctg cgtcatagcc aaactcaaca agggcattac cataaaaaaa 39660  
tttcaattct ggatacgtag atgattttatt tgttacagta ttttttttgg ttttaaccat 39720  
aaccatgta tgtaaatcat tgtagtcagc aacgtaagga gcaaatccaa gttgaatata 39780  
ttcttcttta tagtctttgc taaatttact gttaatatcc tttaacccaa agccccgata 39840  
ccacttccgg gtttcccagt ctccatggat accataaaaa ttatttttaa tacttccagg 39900  
cacaattcct gattgaaaat ataaattcct ttgggatgta aaagtatttt ttctattcac 39960  
caaataagta aaacgaaaaat agtttttctgt gttcttataa aatttatttt tagccctttc 40020  
taggccaatg gaatatattgt aggttgagga gtagtgaaa taaatagaat cattaaaagg 40080  
atctgattta tacattatgg ttgttccgcc agagtaggat atcgggtcttg catcagcatt 40140  
aaagaccatt aaaaaaacta aaatgattat ttatatata cccatggggg gtatcttacc 40200  
ataggattat aaattaattt caggcatctc tccaagcgc cattttcaat gaaatcacca 40260  
tcagaattcc aaaaaataacc atagctatag cagatgcata aaaaatacta actagcccat 40320  
atgagagcca gtaaacagct accaatgcta taacagaaaa tactataagc ctgatggccg 40380  
tgattataat tggccatttc attgcattgg ctccctgaga tgcaaaaatac agggaaaagg 40440

09847513.080801



cataccttg aaatacataa catacaccaa ggatttgaat atactgcttt gtaactaata 40500  
aggtctctgg atccgatga aaaatactaa tccaaagatt tgggtgtagt gctaaaagcta 40560  
gaccgataac agccgaaagc aatccagctg ttgttgcacc aaccatgcct attttttctg 40620  
ctcgtctat atttttggca cctatatattg ttccaaccat ggctgtcata gcagtcoccaa 40680  
tgccgaacac aattgggatt aataagaact ctacccttga accaattcca taaccggcaa 40740  
tagcagacgt tccaaattga ccaatgagtc ctgtcagtaa aagaactgtt gcaactgtca 40800  
ttattggcga aagagatgcg ggtaaggcaa cagaaaaaat atcttcgaat aattctcttt 40860  
caattgttag tcttttgagc ttaagcctta caggcgatga agggctcgaa attttttatta 40920  
aagtactaa tgccataaaa cctgaagtaa ctatcattga ccaagcagat ccacacaagcc 40980  
caagtttagg tagaccgaat gatccctaata taaatccagc agagaaaaaa acttgtatatac 41040  
ctgcacaaat taccgttagg actgcaggaa attgcatac tccatccct cttagggcgg 41100  
cggttaagct acctgatagc caaacaacta ttgctcccaa taaatagaca aaacaataag 41160  
ctaaagactc ctctaataat gcacctgttc cacctaatat tcttaataga ggctctccaa 41220  
aaataaaaaa gataattaaa aaagctaaaag ctccaaagca agatatatac aaagaatgcc 41280  
ataataattt ttcagcccta ggcttgtctg cggctccaag acttcttgca atagaagatg 41340  
tgaccgctcc tccaagagca ccaaatgcca tctgctgagt cagcatgatt gcaggaaatg 41400  
ccaaagtaac agcagctaataa ggagttatgc caagttgact aataaaccag aattcagcaa 41460  
gaacaaccac agcattttatt aaaaatgcaa cagtattagg tgctgacatt ttaatcagca 41520

|             |             |            |             |             |             |       |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| aaggaaatat  | tggatctttt  | aaaaatgct  | ctgttctctt  | gtccataacc  | ttctcattta  | 41580 |
| ttaatattta  | tttggttaa   | gggtgaaaga | gaacacgttc  | tcaggattca  | cgaatcttta  | 41640 |
| aataataata  | acagctttta  | gttaatttaa | tttttaaaaa  | caaccaaatac | tattatctgc  | 41700 |
| tagattttat  | tttttaaca   | tattaatatc | taatttaatt  | gcaaaggggt  | ggccatttgc  | 41760 |
| gcctgagatc  | cttggaacct  | ctttgaacct | gatccataca  | ataatggcgg  | aggaattgca  | 41820 |
| tgaaaaatcat | taaaaatcat  | ttattttatt | caataatctt  | actatatatt  | tttagtctag  | 41880 |
| agacgagctc  | tcaagctata  | gaagaagtta | ttatcaaagg  | agactggaga  | gaaactagtc  | 41940 |
| tgtcagcaga  | agactcaagc  | attgcagtcc | ttgatagtaa  | attaatagaa  | tcccaggctt  | 42000 |
| taaagcactt  | tgaaaaacctt | tcgtacctag | taccaaaattt | aaattttgct  | gcaagtgatt  | 42060 |
| ctagagcaag  | acattttccag | ataagaggaa | ttggagaaaag | atctgggctat | gaaagaactc  | 42120 |
| caaactctgc  | agtaggttta  | ctgattgatg | atatagactt  | ttctgggcaa  | ggtggaattg  | 42180 |
| ccacaacttt  | tgatgttgat  | caaattgaag | ttcacagagg  | cccccaagga  | gcaagaatatg | 42240 |
| ggtctagcgc  | aatggcgggg  | ctcatatata | tttctacaaa  | agacccaaca  | gaaagctttg  | 42300 |
| aagggaaggg  | tgaaatagtt  | atggggtcat | atggaaacttt | taataccgga  | attgctgttg  | 42360 |
| gcggtcctgt  | aaatttttaat | aaagacctca | cttatagggt  | agctataaaag | aaagatttatt | 42420 |
| ctgacgggatt | taggaagaat  | atttttctca | acaagtctga  | tacttctaag  | aaagatgaga  | 42480 |
| gcacatttag  | attaaaaagt  | aattgggtaa | ctgataacca  | aaccacttat  | aagttttctta | 42540 |

|              |             |             |             |             |             |       |
|--------------|-------------|-------------|-------------|-------------|-------------|-------|
| tatctcaaat   | agagttagat  | gacctgcag   | atatatggac  | tatagatggg  | agtcttaata  | 42600 |
| cattatctga   | taagcctgga  | atggactctc  | aaaaaagtaa  | tgcctatggg  | gtaaaaaattt | 42660 |
| atcatcaatt   | taaaaaaattt | gaatttcaaa  | gcctatcaag  | cttaacgaat  | actgatgtca  | 42720 |
| ttcttagtta   | tgatgctgat  | tggggcaacc  | ctgagtcgca  | ttcacccctt  | atttatgact  | 42780 |
| atttttcaga   | aactacaaga  | aaaagagata  | cttttagtca  | agaatttaga  | cttgtatccc  | 42840 |
| aatttgcaga   | taaaaaataca | gaaaaaaagca | tcgaatgggt  | cgttgggggt  | agtttttgtag | 42900 |
| atataaatga   | aacaaatgct  | aaaaaaagata | ctggtattta  | tggagatcca  | tcagaccccat | 42960 |
| atggtcctta   | tttttagtaac | tcttcttctt  | tgagtgactt  | ttcttcttca  | agctatttctt | 43020 |
| tatttgaaa    | tattgattat  | ttaattaatg  | aaacaataaa  | aatttcaatt  | ggtggaagat  | 43080 |
| gggagaattt   | taaatctaata | tatttttgatt | cttatgacga  | atcatttttca | ccatcaaaata | 43140 |
| aaatgtctgg   | tggtaagtgg  | tcactagtta  | aaactcttaa  | taataattct  | aataatttatt | 43200 |
| ttaatatattgc | taagggtctat | aaccagggtg  | gattttaaatt | aggtccttgg  | cttgataaaaa | 43260 |
| attcatcaaa   | tagaaaattta | tattatgac   | cagaattttt  | aactaattac  | gaagtttgga  | 43320 |
| ttaatagtaa   | atttttccag  | tcaaaaattaa | atcttggagc  | agtcctgttt  | tattctgac   | 43380 |
| gaaaagatca   | acaggtctta  | atttcaaccc  | aggttgatcc  | ttcagatccc  | aacactttttt | 43440 |
| tatacttaac   | ccaaaatgct  | gcagaaaggaa | ttaataatgg  | cttagagtta  | aatatagatt  | 43500 |
| atgcactgaa   | taaatctcta  | ggtatatattg | ctaatttttg  | attgctaaat  | acagaaaataa | 43560 |
| aaaattggat   | ttcaagacca  | gatatagaag  | gtagagaaca  | ggcgcgtgct  | ccaaaaaata  | 43620 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| gtttttcaat  | aggcatcaat  | tggaagccaa  | caaaccaatc  | ttatttatca  | ttgaatgttg  | 43680 |
| ttggtaaaag  | tgagtttctat | tactctgatt  | cccataacaa  | tacttctgag  | tcatacaact  | 43740 |
| taacaaacat  | taattatgga  | tatgaacatg  | gacaatggac  | ttattcatta  | tgggcaagaa  | 43800 |
| atattttttga | taagtactat  | tcagtaagag  | gttttttactt | tggtaacgag  | gcacctgatt  | 43860 |
| ttatagatac  | gctctacaga  | agacatggag  | atccaaggca  | tataggagtc  | atgggtccaat | 43920 |
| atgattttcta | actttataag  | tgaattttgg  | atggaaatag  | cagctgttgt  | atttgctatt  | 43980 |
| atttattttat | tacttgctgt  | aaaacaagac  | gtaaagtgtt  | ggtttgcggc  | tataattagc  | 44040 |
| tcaatattat  | atttctttat  | tatgtatgac  | gcaggcttat  | atatggaagc  | ttacctgcaa  | 44100 |
| atttttttata | tcatgatggc  | cttttatggg  | cttcaacagt  | ggagaagtgt  | cgatactgat  | 44160 |
| gctcctcaat  | ttattgttag  | aacatgggat  | aaaaggatgc  | atattaagat  | cattacatta  | 44220 |
| atagttatca  | tgacttttaat | ctctggattt  | ttattagaaa  | aataataccaa | tgctatcttg  | 44280 |
| cctttcatag  | atggactaac  | aacatgggga  | gcaatagttg  | cgacatatat  | ggttgccaaa  | 44340 |
| agacttcttg  | aaaattggat  | atattgggtt  | gttatagatt  | ttatttctat  | cttttttattc | 44400 |
| atgtccagag  | gactttttatt | aacctcagga  | ttatttcttta | tttatcttgt  | aataatatat  | 44460 |
| tttggttaca  | tgtcatgggt  | aaaaataaga  | gacgatatta  | gtgcagaatc  | ttcataacaa  | 44520 |
| tctagatata  | aaaaatcata  | atctagagat  | tattaaaaca  | attaaatcag  | gccctgtatc  | 44580 |
| tgaataatatc | atttgttaatt | ttgataaacat | caaagcaatc  | ttaagagttg  | atcatccatg  | 44640 |

|            |             |            |             |             |            |       |
|------------|-------------|------------|-------------|-------------|------------|-------|
| cgcacacaaa | ataaatgtag  | atcgagaaaa | tgaaatTTTT  | atactcagcc  | aactaaaaat | 44700 |
| tttagatttt | agtcagagg   | ttttatttag | tgatttgtct  | tatggaattt  | tagtatggag | 44760 |
| atatatcgag | ggtattgaat  | tttcacttgg | taaggattcc  | aatgaagttt  | ttttaaaaac | 44820 |
| acttggaaac | gaattaaaaa  | aaattcatga | tattgatctc  | cccaaaaagta | agaaaaaata | 44880 |
| ttttagcaat | gacataaaatt | tttacagaaa | tttactaaag  | gaggttcctg  | aaaatataat | 44940 |
| tcttcataga | ggatttgatt  | tatacgacaa | actcaataac  | tctgataaatt | atgttctctc | 45000 |
| tcataatgac | ttgaataaga  | caaaccttct | ttggagggat  | aggttatttt  | ttttagactg | 45060 |
| ggagtactcg | agttttaaca  | atcctttttt | tgataattgcc | tcgttatcga  | atgcttataa | 45120 |
| cttatcaaaa | gttgataggg  | caatttttat | gaaagcttat  | acaaataatg  | aatattcagt | 45180 |
| attaaatgat | acaaatctta  | gagaatggat | gcatttttgt  | cattatttag  | agtacatgtg | 45240 |
| gagtatttcg | ctcatacaaa  | atggaaaaat | tgatcagaac  | accctaaatt  | taaaaaaatt | 45300 |
| agagaaaaaa | ttaaaaaata  | ttatttaaac | aacaagtatt  | gcgcattctt  | tagttgttgt | 45360 |
| tttactatta | tgtaagaaat  | ttacatttaa | atgggggttt  | gatatgaaaa  | ttttatgtgt | 45420 |
| cttatatgat | gatccaaaaa  | caggtatgcc | agaaaggat   | gcaagagatg  | atttaccaaa | 45480 |
| gttagataag | tatcctgatg  | gaatgacact | tccatcccca  | aaatctatag  | attttactcc | 45540 |
| tggtgagtta | cttgggttgt  | tatctggaga | actagggctt  | cgaaaagtct  | ttgaagatgc | 45600 |
| tggccataca | ctagttgtta  | cttctgataa | ggatggagat  | ggatgtgagg  | ctgataaaga | 45660 |
| attagtagat | gctgatatgt  | ttatatcaca | acctttcttc  | ccatatattt  | taacaagaga | 45720 |

|  |       |
|--|-------|
| caagatgaaa acagcgccta atttaaaaaat ggcaattaca gctggcattg ggtctgatca | 45780 |
| tgttgatctt caggcagcca tggataatag cgtggatgtt gttgaagtta cttactgtaa  | 45840 |
| ttctcgttca gtcgccgaac acatttgtgat gatgaccta tcgatgggtc gtgattatca  | 45900 |
| taccaacat agaattgtaa aagagggagg atggaatata gctgatgctg tacaaaagtc   | 45960 |
| atatgatgta gagggcatgc atgttgggtac ttagcctgca ggtagaattg gtattgatat | 46020 |
| gctaagaaaa atgaaaccct ttgatgtgca ttgcatcac ttgatattc ataaactctc    | 46080 |
| tgatgaaata gaagcagaac taaacctcac ctatcatgat tctgtagagt cattagttgc  | 46140 |
| tgtgtgtgat gtagttaata ttagttgccc attgcacct aaaactgagc acttattcga   | 46200 |
| tgatgaaatg attagtaaaa tgaaaagagg tgcatatatc atcaatactg ctcgtggcaa  | 46260 |
| gatttgtgat aaagatgcta ttgcaagagg cttagagtca ggccagctaa gtggttatgc  | 46320 |
| tggtgatgtt tggttccac aaccagctcc aaatgatcac gtatggagaa caatgcctaa   | 46380 |
| ccacggcatg actcctcata cttcagggtac ttcgctatct gctcaaaaca gatatgccgc | 46440 |
| tgggggttaga gaaattctag aatgttattt tgcaggcgaa ccaattagag acccataatt | 46500 |
| gattgttcaa aatggtgac ttgcagggtat gggcgccac tcgtacacaa aggggtacagc  | 46560 |
| cacagatggc tcagaagagg ccgctaagta taaaaataa gttttagaac ttacttagcc   | 46620 |
| tttcttaat aatagagtct gcctctgcc aatatgctatg catataattt tccactgtcg   | 46680 |
| ggatgtcatt aaccagtcca gcaaccattc cacatgacca ggctccaacc tccatagttc  | 46740 |

|  |       |
|--|-------|
| cttcatgcat aatttttgga tagactcctg caacctcatc cacaatatca gcaaaagtta    | 46800 |
| attcatcacc aagagctttt tctttttcaa tcaatctctc aacagcttca ttattaagaa    | 46860 |
| ccctttctgt atttggttaat gatctcatga tgagtctagt atctaaactca gaagcattta  | 46920 |
| cgatagcctc tttcacattt tgatgaacag gtgcatcttg agtagcaata aacctagttc    | 46980 |
| ccatattcat tccctcagca cctagtgaca tggcagcaac caaacttctt ccatctgcca    | 47040 |
| ttccacccga ggcaacaaat ggtatttcaa gctcatctgc tgctctaggt aagagtatga    | 47100 |
| aattaggaat gtcattcttct cctgggtgtc cgccacactc aaacctatct acagaaaccg   | 47160 |
| catcgcaacc tattgcttgt gctttttaatg agtgcccttac agaggtgcat ttatgaataa  | 47220 |
| cttttattcc tgcttctttt aaagctggga ggtattctgc aggatttctt ccagcagtct    | 47280 |
| ctacaacagg aacccttgca tcaataatta cttttatcaa accaggatag tctgggggtg    | 47340 |
| ttagtgatgg taaaaatgtt aaattaacag caaatggctt attggtcat tctttgcatc     | 47400 |
| tggcaatttc atttgctaatt ttctcaggcg taccctgtgt tagaccagta attgttccaa   | 47460 |
| gtccaccgcg atttgatact gccgctgcaa gctcgggcaaa accaacatgg tgcattccac   | 47520 |
| cttgaatgat tggatgttct ataccaaata attcagttat tttagttttc ataatactct    | 47580 |
| cctattttatc ttccattggg gttaaaaaaat cttcatattg agtttgaagt ctttgcattgc | 47640 |
| cacttatcca tctatctctg tcattacctt ttcttttaac atatctaga acctcactat     | 47700 |
| gaggggttac taaaaacctc tcctcctcta tggcggttaag aacatccttt gcaacaatat   | 47760 |
| cagcttccat cattccatca actcctgcta caccaggacc atttgcagtc atagcagttc    | 47820 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| taactgcttg  | agggcataaa  | caggaaaacgc | cgataaccttt | atttccatat  | gtaattttta  | 47880 |
| tccattcagc  | aaaactcaca  | gctgcagctt  | ttgtaactgc  | atatccagcc  | gcacctagtt  | 47940 |
| gagttaaaag  | acgggctgct  | gaggatgtat  | tcataagata  | gccttcacct  | tgttctatca  | 48000 |
| tttgaggaag  | cacatgtttt  | gcagcatgaa  | tatgagactg  | aacattaca   | ccccatatca  | 48060 |
| tgtcccaatc  | cgaagtatct  | gcttcaaaaa  | atcctggcctt | tcaccgata   | cctgcatttg  | 48120 |
| aacaaaatat  | atcaatacca  | ccagaaaatt  | cattagcctt  | ttgtataaca  | tttataatgt  | 48180 |
| cgttttcttt  | ggaaaacatca | gcacttactg  | caagcccatt  | aacactcttg  | gctgttttcct | 48240 |
| ccgccccatt  | taaattcata  | tcaacacata  | ctatagaact  | tgccccagat  | gcatagaact  | 48300 |
| cttcacataa  | agccttacca  | attccactgg  | cagcccctgt  | aaccacaaact | cttttattat  | 48360 |
| taattttcat  | aaattgacca  | ctttttttaat | tttttattta  | actttattgc  | ttagatagtt  | 48420 |
| tattacaaaa  | caacaatact  | taatattgca  | aaatcgcata  | cccatcttta  | taaattttatt | 48480 |
| gctattatta  | gtgctagaaa  | atgaaatact  | gagatttaat  | atgaagaatg  | ttgtttgttat | 48540 |
| tggctcgtcc  | ggggcaatag  | gaaaagcctt  | tattgatagc  | tatatcaaag  | atgatgatgt  | 48600 |
| tgaaaaatata | ttttcatttt  | caagaacagg  | cctttccatt  | gaggataaaa  | aactcccatag | 48660 |
| tttttttatt  | gatattgagg  | atgaaaactag | tatttgtgat  | gccgcagaga  | agatagacaa  | 48720 |
| gtcctcaata  | gatgaaatta  | tcgtcgcgaag | tggaatactt  | cataataaag  | attttgggcc  | 48780 |
| agaaaaaaagt | attagagatt  | taaatgcaga  | taacctttta  | aaggtcatta  | aggttaatac  | 48840 |

09847513.080801



|            |             |             |            |            |             |       |
|------------|-------------|-------------|------------|------------|-------------|-------|
| tatcggccca | acaattgttg  | gaaagtattt  | cattccattg | ctaaataaaa | aagaaaaaag  | 48900 |
| cgtcttagca | ttttaagtg   | caagagtcgg  | cagcatttct | gataataaaa | caggtggttg  | 48960 |
| gtatgcctat | agagcgagta  | aaactgcact  | taatcaaadc | attaaaaagt | ttagttattga | 49020 |
| attacgaaga | accaatccaa  | atgccattat  | ttttggtcta | cagccaggaa | cagtagatag  | 49080 |
| tgaattaagc | gaacctttta  | aaagaaatgt  | aaaagaaggt | aatttattta | ctccagaata  | 49140 |
| tagtgtattg | cagctaaaaa  | atattattga  | tacagcaagt | ccatctgatt | caggcaaaact | 49200 |
| aatttcttgg | gatggggaag  | agattcagcc  | atagttggtt | atgaatatat | tttcatatta  | 49260 |
| gaataaaatt | ttaggggaaa  | aaatgagtat  | taaatattat | gactggtcta | aatttcaggc  | 49320 |
| caacactagg | ccaaataaag  | ttgccataag  | agagctagat | aataacaaga | tctatactta  | 49380 |
| cggagaattg | gataaaaagat | catcaaggct  | tgcatacat  | ctccaaagtt | caggaataaa  | 49440 |
| aaaaggagat | cgtattgcga  | tactatcgct  | aaattgttca | gaattttttg | agctagaatt  | 49500 |
| tgcttgcgga | aagattgggg  | caatagagat  | accattaaat | tggagattaa | caaaaaccga  | 49560 |
| gctcagttat | attcttaatg  | atagtgagcc  | aaaaactcta | atttatgaca | atcagtttga  | 49620 |
| agaaatggtg | aaagagctaa  | aagaagaatg  | taatatctct | gaaatcatag | ctcttgatca  | 49680 |
| atttgacca  | gaaagtgatt  | atgaaaaagt  | tttgagtaat | gcttcaggca | tttattatca  | 49740 |
| ggaagaagtt | gatctagaag  | ataacattat  | gattatgtat | acctctggaa | caaccggtca  | 49800 |
| ccctaagggc | gccatgatca  | cgcacaaaat  | gcagcttttt | aatgttatta | atttaggtat  | 49860 |
| ttcagcagct | gtttccccctg | aatcagtgcca | tttagttgtc | cttcctttat | ttcatacagg  | 49920 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| cggaatgaat  | tgttattcaa  | atccaattct  | tcatgcaggt  | ggcgagttaa  | tattacttaa  | 49980 |
| agagtttgag  | cctgggaaag  | ttctatcaat  | tatcggcagc  | tctgactatg  | gagttactca  | 50040 |
| tctgtttgca  | gttccagccc  | cttatcaatt  | tatgatgaat  | catccagatt  | ttgaatcaac  | 50100 |
| aaatttatca  | ggagttaagt  | atgctggagt  | tgggggcgca  | ccttgtgcag  | aggctatttt  | 50160 |
| gaagacttat  | ataagcaagg  | gtgttttcgat | gcagcaagga  | tgggggtatga | cagaaaactag | 50220 |
| tccaggtgct  | actggtcttg  | aatcgtccga  | ggctgaaaga  | aaaataggat  | ctgctggaaa  | 50280 |
| accgcttctt  | catactgagg  | tcaagggtgt  | tggagatgat  | gggaatgaac  | tgcctgctgg  | 50340 |
| agaagtaggc  | gagattttata | ttaaaggccc  | aaatattaca  | cctggctatt  | ggaagaaaaga | 50400 |
| agaggctact  | agagatttctt | ttgaagacgg  | gtggttaaaa  | acaggtgatg  | ctgcttactt  | 50460 |
| tgatgacgag  | ggtttttttat | acatagtttga | tcgatggaag  | gatatgtata  | tctcagggtgg | 50520 |
| agaaaaatgtt | tatccagctg  | aagttsaana  | tgttatctat  | cagttaccac  | aaatcgcaga  | 50580 |
| agttggaggt  | attggtattg  | atagccctaa  | gtgggggtgag | actggtaaag  | cctttgttgc  | 50640 |
| tttaaaagccc | gatcatgaat  | tgacggcaga  | agaagtcata  | gatcattgtt  | taaaaaatct  | 50700 |
| agcaaaagtac | aaaattccag  | agaaaagtga  | gtttatttga  | gctcttccaa  | gaaatgctac  | 50760 |
| aggtaaagtt  | ttaaaaagaa  | cattaagaga  | tatgtaatat  | ttttaagcca  | aaaaaaaaccc | 50820 |
| agcctaagct  | gggttttttta | tttaactatt  | taaagttatc  | taatgttaag  | cattagaaga  | 50880 |
| ttctttaaca  | gcaacattcc  | atataattaa  | acaaaataga  | atcttgttta  | caaagtcagc  | 50940 |

|             |             |            |             |             |             |       |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| aaggttatag  | ataaggttta  | agttaagagc | tgatccaccg  | tcacccatca  | ggtaacctgt  | 51000 |
| gaaataacct  | acaggataaa  | tcgccaacc  | aaagatgata  | atatacatca  | ttgtgttgta  | 51060 |
| agctgattgc  | acagcaggac  | ttgcagtatt | acatgcagat  | tttccttctc  | cagcccataa  | 51120 |
| ttcataaatc  | atgtataccc  | aagctaaaca | ccaataatg   | aatgcaggcc  | atgcagccat  | 51180 |
| gattcctgct  | tcacccatgt  | aaccaaacac | aagcataaca  | agagaaccaa  | ctagtaattt  | 51240 |
| cttaaaataat | gatccagcaa  | cattagttgc | agcagcaaga  | attaagtaga  | attcacatat  | 51300 |
| taatagagga  | actgttagta  | accaatcaat | gtatctaaat  | acagttggcg  | aatcaccagt  | 51360 |
| ttcaatccat  | acccctctca  | tgtacatgta | atgccagaaa  | gcaataccag  | taacaagacc  | 51420 |
| agatacagtt  | aatgatgttt  | tccattttgc | agaaactcta  | tctctttcaa  | caaagaaaaa  | 51480 |
| tacagtagat  | gctaataaag  | cagcagtaac | taaccaaaaa  | gaaacaccag  | tgtaatcact  | 51540 |
| agcatcaagg  | tcaccaccac  | ctgcagcaaa | tgtagggaagt | gcaataaacac | tacctaatat  | 51600 |
| cagtaataat  | ttcatatata  | actcctatat | taatgacaaa  | gaagcataat  | tgccctcccc  | 51660 |
| ctaattaaat  | atgaaaattta | atagtaacca | atattaatag  | aggttacaaa  | taaaaaagcat | 51720 |
| ctattttttt  | aatagaaaaat | atgtgtaata | aaacaccata  | ctttaaatta  | atatttat    | 51780 |
| tgcataagta  | gacttgttta  | gtatcatatt | taacatgaaa  | gttgcaattt  | atcctggttc  | 51840 |
| ctttgacccc  | atcacaaaatg | gtcacacgga | tattattgat  | aggggttgcg  | gactcttttga | 51900 |
| caaggttggt  | gttgcaatag  | ctaagagtga | atcgaaaaaac | cctcttttta  | gcctagagga  | 51960 |
| tagaattaat  | ttagcccaat  | ctatttttta | aggaaatgaa  | aaagtagagg  | ttgttggttt  | 52020 |

|             |             |             |             |             |              |       |
|-------------|-------------|-------------|-------------|-------------|--------------|-------|
| tccaagaaag  | ttaacagttg  | atcttgcaaa  | agactatgga  | gcttgtgcaa  | ttataagagg   | 52080 |
| cctacgagca  | gtttctgatt  | ttgaatatga  | atttcagtta  | gcaacaatga  | ataggtcgc    | 52140 |
| ggctcccaat  | attgaaaagca | tttttttaac  | accaaaagaa  | agtcctcatt  | atgtatcttc   | 52200 |
| tagcttaatt  | aaagaaatat  | cagacttaaa  | aggcgatata  | tcgaagtttg  | ttcatcctat   | 52260 |
| agtagagcag  | gcacttcgag  | cgagtagaca  | cttagctctg  | acaagcttca  | caaaaaaaaaag | 52320 |
| tagctctttg  | atgacgac    | gtttacaaa   | tagtgccttt  | gcatttattg  | cagggctcgc   | 52380 |
| cttctcttcc  | atatacattt  | aattttaaatt | taaaatatcc  | tggactacca  | tcagctgagt   | 52440 |
| agaaagtcttt | taatgtagtt  | cctccaaacct | ctattgcttg  | ctctagtatc  | ttttttccag   | 52500 |
| cagctactaa  | tcttttacaa  | gcactctaaat | ctaactcatt  | ggcatttttt  | agcgggatgaa  | 52560 |
| tttttagctaa | gaaaaggcctt | tcagacgcac  | aaatattgcc  | gataccgact  | acatttttttt  | 52620 |
| gattcattaa  | atagctttta  | atatttgttt  | tagagtgact  | gcacacctgaa | aaaaaaatcct  | 52680 |
| ttgcattaaa  | attttttagaa | agaggttcag  | gcccaagggt  | ttttattaac  | ttatgttttat  | 52740 |
| cgatatcaga  | agtaagatgc  | attgatccaa  | accttctggg  | atcattataa  | ataaattcttt  | 52800 |
| cctcatcaaa  | aattaattca  | atatgatcat  | gtttgataaa  | gaagtttttca | ttatttttttg  | 52860 |
| caattctaag  | actcccagac  | attcctaaat  | gaagaataat  | tttttttatca | ttagatagtt   | 52920 |
| taaaaaattat | atatttttgct | cttcttttcta | agcttttcgac | aacttgatct  | tttactgagg   | 52980 |
| tttcaaaagct | atccaccact  | ttccatctaa  | ggtttcttgtt | gtgtatcctt  | gcttcttttca  | 53040 |

09847513 . 080801

|             |             |            |             |             |              |            |       |       |
|-------------|-------------|------------|-------------|-------------|--------------|------------|-------|-------|
| gcaaagagcc  | tttaaat     | ttt        | ttaattgctc  | tcaggggtgt  | ttcaacctct   | ggaagt     | ttcag | 53100 |
| gcatttttaa  | ctgagtaaat  | tattaatttt | tactatgtca  | gctggagtaa  | ttgttcccga   |            |       | 53160 |
| tgcgagccct  | agtcttagat  | tggcaagtat | gtaatcatat  | tttgcattag  | caagattttt   |            |       | 53220 |
| ttccgcgctg  | tataagt     | ttt        | tctgcctg    | caagagatca  | acaacgtttc   | ttgttccaac |       | 53280 |
| tctgtagcca  | acttgagtcg  | cttccagggc | actagtggcc  | gaaatcactg  | cttgttttttg  |            |       | 53340 |
| agcattttaca | tttgcaacta  | atgttaaaac | atttgaaaac  | tgggatctga  | cttctttgaat  |            |       | 53400 |
| aatccttctt  | tctgtaaata  | gagtattttc | atttgctctt  | tcatactgtg  | aatatgcttg   |            |       | 53460 |
| cttccttctt  | gagttaacgg  | cgccgccttg | aaagagtggc  | atacttagct  | gaattgcata   |            |       | 53520 |
| attccttctt  | cctgttactg  | atggaactgg | aataccttgg  | ccattgatat  | taaaaccttc   |            |       | 53580 |
| atagttaaat  | tggtttg     | ttt        | cagatttctga | ctgacttcca  | acaatgtcta   | tcttaggtaa |       | 53640 |
| atgatttgaa  | gctacacttc  | ttgcactgct | tttcgctgct  | ttcttttctca | aatatgctgc   |            |       | 53700 |
| ttttaactgg  | tagttatttt  | ccaatgctaa | ttcaacccat  | gtctcttttg  | aacttggtgt   |            |       | 53760 |
| tggcagggtca | ataagcaaac  | catctcccaa | ttcattttaag | ctgaatatatt | ctctaccaat   |            |       | 53820 |
| cagagcattt  | aaagactctc  | ttgcagaata | aagtgatcct  | tctgttctaa  | ttcttgaggc   |            |       | 53880 |
| tttactttaga | tcaaatgccca | attgagcctc | ttgaactcca  | gttatggctg  | ataacccaac   |            |       | 53940 |
| atcgaatctt  | tgttttgctt  | gatcaagttg | ttttttaata  | gctttttctt  | cagatatattgc |            |       | 54000 |
| tgcattttaga | ttatcaatag  | ctctaagtac | gccaaaaataa | agctcagcag  | ttcttactaa   |            |       | 54060 |
| aagatttttgc | tgtctaaaatg | caaagtctgc | ttcagcagca  | tctgtaagag  | atttagattg   |            |       | 54120 |

cctatatga aaccatgtat ctagtctaaa gagcggctga gtaaccctgg cagatgtaga 54180  
aaaagagtta tattgctgct gcagttcttt gttttgatag tattcgttcc agttagttga 54240  
tccactcaag gtaatactcg ggagaaagcg agctcttctt tgaacccttaa gctcttttatc 54300  
tgctaaatat gaatattccg ctgctttata tgtagggtca ttctcaagcg cttcattata 54360  
gatatccaaa agacttttcag atgaaatatt aaatgagata aaaagtgcta aaaatgatatt 54420  
tgtgtaaaatt ttcataaactt attttaacct attaatgaatg ttgcagtcg aaacatttat 54480  
ttttttttaat tttattgaat ttattttata tttcttagag tagaatatct ctaacaagtt 54540  
caataaatttt ttataataaa aaggcttttaa aaattggcta aaaattcata tgacgctcag 54600  
gcaattgaag tcttatctgg actagatcct gtcaaaaaaa gacctgggat gtatacggat 54660  
acatctaacc caaatcattt aattcaagaa gttcttgata attcgggtga tgaagctctt 54720  
tcagggtatt gttcaaacat aaaagtatct gttctaaaaag atggctttat taaggctctct 54780  
gatgatggaa ggggtatgcc aattgatgag caccgggaac ataaagtttc aggtgttgag 54840  
ctcatccttt gtaaaacttca tgctggagcg aaattctctg gagatgatta taatttctct 54900  
ggtggccttc atggtgttgg agtttctgtt gtaaatgccc tatctgatga attagaggtt 54960  
agagtaaaaa gagattctaa agaataccaa atcactttta ataatggaga taagtcttct 55020  
gaattaaagc caattgggga agtggggctc agaaattctg ggacatcaat taaattcaaa 55080  
ccaaatccta catattttga aactatagag attcagataa aacagcttaa gcattttatta 55140

aaggccaaag cagttctctg tcctggatta acgatagagt tcgttaatga aaaaaagact 55200  
gatgataaac aaagtggta ttttgaagat gggctcaaaa gctatttgat tgattcttcc 55260  
gagggagcag acttggtttt gctagattca attgtatgct ctaaaaaatc tcaagctcaa 55320  
gagcttgaat ttgcaatcaa ttggtcatta agaccccca aaaataaaact cgatgaaacc 55380  
tatgtgaatc tcataccaac tgctcagggt ggctcacatt taaatggctt taaggctggg 55440  
cttttagatt catataaaga attttgtgaa tacagaaaac tattgcctaa aggttttaaa 55500  
attaatgcag atgatgttct taataatgca atttttataa ttcatctaa gcttcagaat 55560  
cctcaatttg cagggcaaac caaggaaaga ctagattcaa aagatcacat gtcgttcgtc 55620  
tcaagtacca caaagacat tttaagtatt tggcttaaca ctcatacaga agagggcgaa 55680  
agaatagcag aacttgcaat tatgtctgct cagacgagag caaaagtctc caatatagtt 55740  
gaaagaaaga aaacttttag aggccagcc ttacctgga aactttcgga ctgtaatagt 55800  
caggacttaa atgaaacaga gcttttttta gttgaggggg actcagccgg agggtcgca 55860  
aaacaagcaa gagaaagatc tttccaggca atcatgcctt tgagagggaa gattttaaat 55920  
acttgggact tagaaagtgc agaaataata aaatctcagg agataaaaaa cctatcaact 55980  
gcaattgggg ttctgccagg aaataatgac ctttcatcac taagatacgg aaaaatttgt 56040  
attcttgcag atgctgattc agatggtctg catattgcaa ctttactttg tgcattgttt 56100  
ctaaggcatt ataaatcttt agttcaagag ggaaggatat atatttcaat gcctcctcta 56160  
tatagaatcg attctggtaa agatgttcta tatgcacttg atgataaaca gcgagatgaa 56220

|  |       |
|--|-------|
| atagttactg aatttaaaa gaagaagggc aagcctaaag taaacattca aaggtttaaa   | 56280 |
| ggacttggtg aaatgaatcc acctcaacta agagagactg tgatggaccc tgctactcgt  | 56340 |
| cagcttggtc agctttctat cagctcaagc gataatgcaa attctatgat ggacttactt  | 56400 |
| ttgtccaaaa agaacgcacc agcaagaaaa gaatggcttg aaaagaaaagg gtctctagca | 56460 |
| aaaatataaa tatgaaagaa caaataaacct caattagcct caagcaatat gctgaagagt | 56520 |
| cttatcttaa ctatgcaatg tatgtcattt tagatagagc ttgcctaata attggagatg  | 56580 |
| gccttaagcc tgttcaaaga agaataactct atgcaatgtc agagcttggg cttgatgctg | 56640 |
| gctcaaagta caaaaaatca gcaagaactg ttggagatgt tataaggaaaa ttcatcctcc | 56700 |
| atggagacgg cgctgcatat gaagctatgg ttttaatggc tcaaaatttc tcattcaaat  | 56760 |
| accttttgt agatggtcaa ggtaactggg gttctcagga tgatccaaaa tcttttgctg   | 56820 |
| caatgaggta tacagaaatct aagttaacta aatttgcaaa tcttttaatc tctgaattga | 56880 |
| agtctggaac agtcgattgg cagcctaatt ttgatggctc tcttttagag ccagtaattt  | 56940 |
| ttccagccaa actcccatct atttattga atggcacttc tggaattgct gtaggaatgg   | 57000 |
| caacagatat tccatctcat aatattaatg aaattattga tgccacagta catcttattg  | 57060 |
| ataatccaaa atcacagttg gttgatcttac tcaagataat taatggctct gatttctcaa | 57120 |
| ataattcgcc aataattgct agcaaagatg agctgaatga aatttattcg actggaaaaa  | 57180 |
| gcggcttcaa agctcaagcc caatggggcg aggataagaa tcaaatcatt atcaacgc    | 57240 |



|             |            |             |             |            |             |       |
|-------------|------------|-------------|-------------|------------|-------------|-------|
| taccttatca  | agcatctggg | tctaaaaattt | tagagcaaat  | agctgatcaa | atgcttaaga  | 57300 |
| aaaaaattcc  | aatggtggtt | gatcttactg  | atgaaggaga  | ccacaaggag | ccagtaaggc  | 57360 |
| ttgtcataac  | tttaaaatcg | aacagagtaa  | atgctgaaga  | tgtaatgaat | cacctttttg  | 57420 |
| catcaactga  | tttacaaaa  | aattatagag  | taaatatgaa  | tttgatttca | ttgaaaagtg  | 57480 |
| gacccaaaagt | tttctcccta | gttgatttat  | tgaagaatg   | gctagtcttt | agaaaaagaaa | 57540 |
| ctgtaataag  | aaaactagaa | catagactcg  | accaggtaaa  | cgataggctg | catatccttg  | 57600 |
| aggggttatt  | aattgtttat | ttagatttag  | ataaagtaat  | aaagattatt | agagaatcag  | 57660 |
| atgagccaaa  | aaaagatat  | attacagctt  | ttaaaatttc  | tgatatccag | gcaaatgcaa  | 57720 |
| tccttgagat  | taggttaagg | caactagcca  | agctagaaca  | aatcaagtta | gaacaaagaaa | 57780 |
| gagatacttt  | ggtagcagaa | caaacagata  | ttgaaaaaat  | actaagctct | aagacaagac  | 57840 |
| ttaaaaacatt | aattaaaaat | gagcttat    | tg          | tg         | tg          | 57900 |
| aatctccaat  | aagagaagca | acagaagcca  | aagttttttc  | tgaagaagaa | actcttgtca  | 57960 |
| ctgagcctat  | tacagtagtc | ttgtctgcag  | ctggctggat  | cagaagtgc  | aaaggccatg  | 58020 |
| agatagaccc  | cagctcactt | tcctatagag  | gagaagatgt  | acttcaagat | tatggaagag  | 58080 |
| gaaagagcaa  | tcaagtttca | gttttttcttg | actcaaatgg  | gaaggcttac | tcacttgcaa  | 58140 |
| gtcactctct  | tccatctgct | agaggaatgg  | gtgatacctat | tacaggaagg | gtatctgcag  | 58200 |
| atttctggagt | aaagttcatt | tcttcattga  | ttgggaatga  | tgaagataaa | ttcatgatta  | 58260 |
| tgaataactgc | tggatatggc | tatatcttcag | agtttaaaaa  | tatggtttct | aataaagaaat | 58320 |

caggaaaaagc atttatgaaa atccctcatg aagcagacct tcttaaagcc attaaaagtaa 58380  
gagacgatca tttgtatata gcagcagttt caaatattgg cagactttta attttttaaga 58440  
ttgatgaatt gccaaactctt ggaaagggca aagggaataa aataataaat ataccaaccg 58500  
ctaaatttat agcaaaaagaa gagttaatga cccatgcgca acttgtttct gaggctagct 58560  
ctttaaggat tgaaaagcga aagagattcc tcactttaaa actcaaaagat ctagaaaaact 58620  
atatttctac aagagcaaaa aggggaaata tgcttccaca aggatattag aaggtagata 58680  
aaatgattga agagggttgag ttagaagtta aagaagactg attatagatt tctcaaaaaat 58740  
ttttaatccc tcatcaacta gctcatttct tattattagg ctggagaaa acctaacagt 58800  
agatgcattc gccttaagaa tcattaaccc attatgatgg gattttttta ttaaatcattc 58860  
tatttgaatt ttgctatcct tgcttacttc cactccgacc cataaaccag cagaagttat 58920  
tttttcaaaa catttgtgct tctcatttat ttatttcaac aaattgataa atctaacttc 58980  
tttctttaaa accttattta aaaatgattt tttagaaatg gtatctataa cttcattacc 59040  
tatggcacat gcaattgggc ctctccaaa agtcgtgcca tgagaccctg ccgacatatg 59100  
ttttgaaact ttatctgaag taagaatgcc tcctagtggg aagccattag atataccttt 59160  
agcaaaaaca agaatatcag gagtaatgtt aaattgctca taagcaaaaga gtgttccagt 59220  
cctgccaatc ccggattgaa cttcatcaat tatgacaagt gccttatgtt tttttgcaag 59280  
tttctttatt ttgctataa attttttctc tgctttttgtt attccagatt gccattgaaac 59340

|              |            |             |             |             |             |       |
|--------------|------------|-------------|-------------|-------------|-------------|-------|
| cagttctaataa | ataaccgctg | cagtctttatc | tgaaaaaaac  | ttttctagat  | tagtaatgtc  | 59400 |
| attatatgga   | tgatttttta | tccctctagg  | cagaggcgca  | aaaccgtctg  | taagatgctt  | 59460 |
| agctttagct   | aaagctatac | caagcatagt  | tcttccatga  | aatgacgttg  | aaaatgaaat  | 59520 |
| gacctcattt   | ttatttttat | taactgttga  | tgagcaaaat  | tttcttgcta  | tcttaactgc  | 59580 |
| agcctcaatt   | gattctgccc | ctgaattgca  | aaaaataact  | ttatctgcaa  | aagaatttctt | 59640 |
| gcataacttt   | cttgccaaag | taactgaggg  | ctcattttatg | tagagatttg  | ataaatgccca | 59700 |
| aagctcctca   | gattgctttt | taagaatttt  | tattagatct  | ttatttgaat  | gaccaagggtt | 59760 |
| agtgacggca   | atgcctgctg | tgaaatcaat  | atatttctta  | ttatttaggt  | cccatatcatg | 59820 |
| cgatccactc   | gcttttttta | caacaaaatc  | tgcaggagcg  | taaaaaggca  | ccatataaatt | 59880 |
| agtataattct  | ttaattatct | ctttcataaa  | tatatgtgat  | cgtggtttca  | gttcaaaaaat | 59940 |
| taaatgacca   | aaaatttcta | gtatctttta  | aaccaaacag  | ctcactcatt  | ggctttaata  | 60000 |
| gaataaatatt  | tatatctagt | atatctattg  | tatgcggtgg  | catagcattg  | atattttttct | 60060 |
| ttttttggagc  | aacgctcatt | ttgccttttg  | ctggtttaga  | gcttggtatt  | ctattcactg  | 60120 |
| cattctattt   | aagttttaaa | tggagtgata  | aaaaagaaaa  | aatatattatt | tctcaagatc  | 60180 |
| ttgtaactat   | agaaaaaggc | tctaattatg  | ctgaatatata | atgggaagag  | ttcagggtcat | 60240 |
| ttacctcttt   | ccaggtttca | aaagacagaa  | gagatcttct  | taagctaagc  | tttaggtcta  | 60300 |
| aaggcgagga   | tgttgagggt | ggcagctttc  | taaatgaaga  | tgataaaaaat | gtattaatag  | 60360 |
| aagagctaac   | tcagatcata | gatacattaa  | atcacgattc  | cttctcaaag  | ccagagcctt  | 60420 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| aatttctcttt | taacttcttt  | tagctttaat  | tcttttattt  | cagggatgcc  | atttctaaaa  | 60480 |
| tttggaatat  | atttctccctg | tattaatggc  | ataagatatg  | caattgcttt  | tggggtaacg  | 60540 |
| tccattccat  | tttttagctat | aaaagatttt  | ggtaattttt  | tttctaaaatt | agctattttt  | 60600 |
| gccagtggag  | ctggttcaat  | tttccaccta  | tatttttttag | ctttacctct  | aactatttatg | 60660 |
| ggcataaacac | cattcatgcc  | tttctactgca | tactgaactg  | ccttggcccc  | aacagccata  | 60720 |
| gcttgctcta  | ggtctgtttt  | ggaggcaata  | tgctgtgcac  | ttctttgtag  | atagtcagaa  | 60780 |
| acagcccaat  | gatttttttag | ttttaattta  | tcagtaatta  | agttggcaat  | gtaaggcgca  | 60840 |
| actccacca   | gttgagcatg  | accaaatgca  | tctattgttg  | ctgattcaga  | aagaaaatctg | 60900 |
| ttcttattat  | ttttcaaccc  | ttcggatacc  | acaacaacac  | agtagccatt  | ttttttaaca  | 60960 |
| acacttttaa  | cttctgctag  | aaattttttt  | tgattgaagg  | ttatttcagg  | taataaaaatt | 61020 |
| atatgagggtg | catctccttt  | ttcttttctt  | gcaagagaaag | atgcagccgc  | catccagcca  | 61080 |
| gcatgcctgc  | ccataacttc  | taataataaaa | actttttgttg | atgttgctgc  | cattgaccga  | 61140 |
| acatctaagt  | aggcttcttag | agttgatgtt  | gctatatatt  | tagccgccga  | cccaaatcca  | 61200 |
| ggacaacagt  | ctgttaatac  | caaatcgtta  | tcaacagttt  | ttgggatggc  | tatacagtta  | 61260 |
| attggataat  | ttaatttttt  | acttatctgc  | gaaacttta   | atgcagtatc  | agctgaatca  | 61320 |
| ttaccgccat  | tgtaaaaaaa  | atatacctata | ttgtgcgcctt | taaaaacatc  | aataagcctt  | 61380 |
| ttgtactctc  | tttcgcttga  | ctcaagatct  | tttaacttaa  | acctacatga  | cccaaaaagcc | 61440 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| ccaccaggcc  | tataatttcaa | tgactctaat  | gcagatatatg | attcttttga  | tgtatctatt  | 61500 |
| agttcctcat  | tcagecgcgc  | caatattcca  | ttcttgcctg  | cataaaatttt | accgatgtct  | 61560 |
| ttatgcttct  | tggcctctaa  | aatcagagcg  | ccagctgtag  | cattttataac | tgcgggtaacg | 61620 |
| ccgcgggatt  | gagcataaaa  | cgcatttttt  | ttcatTTTTT  | ctccatcaaa  | tgtactgaaa  | 61680 |
| taaaaaatatg | taaagtatta  | taactaatat  | gcgaatacat  | atcttaggga  | tctgtggrac  | 61740 |
| ctttatgggc  | ggccttgcta  | agatacttaa  | agagtcaggg  | catgaaaatat | ctggatcaga  | 61800 |
| tattcaattt  | tatcctccca  | gttcagacta  | tcttgatagc  | tttgatattg  | aaatgatcaa  | 61860 |
| gggctatgac  | ataaaaaagca | tgccagatgc  | tgacttgtat  | gtgattggta  | atgcactttc  | 61920 |
| tagaggaaat  | gaatccgttg  | agcatatattt | gtccaattct  | cttccgttta  | agtcaggccc  | 61980 |
| tgaaatgctt  | ggagaaaattt | taaaaaataa  | aaagggtttta | gctatatctg  | gaacacatgg  | 62040 |
| taaaacaacg  | acttcatata  | tgttgactca  | tattatgtta  | gatcagggta  | gggatgtcgg  | 62100 |
| tttttctagt  | ggtggaatat  | caaataatat  | atctggctct  | gcatgtcttg  | ggtcagatgg  | 62160 |
| aacttttgtg  | attgaggcag  | atgaatatga  | ttctgcattt  | tttgataaaa  | gatcaaaaatt | 62220 |
| tattcactat  | tcaccaagca  | ctatagtcat  | caataatatc  | gaatttgatc  | atgctgatat  | 62280 |
| tttttaataat | ttagatgaca  | ttaaaagaca  | attccatcat  | ctaattaaaa  | taattccatc  | 62340 |
| aaatggaaat  | gttgttttatt | ttgctgatga  | taaaaatata  | agagatctta  | tcgatatggg  | 62400 |
| aatttgggtca | aatcagatag  | caatcaataa  | taatgctcat  | tcaattgagg  | cagttttattc | 62460 |
| tgataagact  | ctaaaaatatg | aagaaagtat  | ttattcatta  | aatgagttac  | ctttaatatg  | 62520 |

agaacataat tttaaaaact acatttcggc tattttggcg gcaaaagacag atggaattca 62580  
 aattcaggat tctatcaaat cattagctag ctttgatggc gtaaaaagaa gattagaata 62640  
 taaagggaagt ttgatggca tagaaattta tgatgatttt gtcatcatc caaccgcaat 62700  
 agaattttct tctaattgcc tcgtaacaca aaatcccatca aaaaaaatatc ttggtctcat 62760  
 tgagctaggc tcaaacacta tgtctggggg ttctcatggc ttgtcttttg ttgaatctgc 62820  
 aaaatcttta actcatgtta tctgggctaga tcgcaataat gttttgtcag agaattgctag 62880  
 cattgaatct actaacacta ctgaagattt tattttctgca gcgatatctg cttttctcaga 62940  
 ttatgatatt gttattttga tgaccaataa agacagccaa aaaatatata aaccattgt 63000  
 agatcacttt gaaaaataat aatttaccag tttttccttt aggaatagtc gccctcccag 63060  
 gtagcatcca atctcttcaa attttttgagc ctaggatat acagatgggtt aaaacatgtc 63120  
 tatccaagaa ccatggattt gtaattgttt ttaatgccaa taatgagtct caaggcgatt 63180  
 tcaacttttc taagaaggga agttttgttg aaattataga ctttaataat ttgccaaatg 63240  
 gccttcttgg gataactgta aaatctataa ataagggtgat aatcagtaat atatgtcaat 63300  
 tagaagatgg actgcatatt gctgatatta aggcacagat agatccagag gtagatgac 63360  
 aagctgtttt ggcagaatat cctgagatat ctagcattct ttctcagctt gtaaagcatc 63420  
 ctaagattag tgacctgcc atccagggtg actttggctc tgctgattca gttgcatacc 63480  
 acttagcagg ccttatacct ctaagctcaa atgagaaaaca aaaactatta gaagcattcg 63540

|             |            |             |              |             |              |       |
|-------------|------------|-------------|--------------|-------------|--------------|-------|
| atgcagcaca  | gcggatgaga | attcttttcag | actatatattga | aagaatatct  | actacataaa   | 63600 |
| ttatttattt  | taatattatt | ggcggcttat  | tagattttgc   | tctaagccaa  | ttgatggact   | 63660 |
| taaaaaccgt  | agggatttct | acaaaccttt  | tttcaatata   | gtatttgcct  | ggatagttac   | 63720 |
| taagcattag  | ccaataata  | atagtgaaca  | aacctggcca   | ggtaacacca  | gcacatgat    | 63780 |
| tcccccaata  | ataagggaat | aacctaaaaat | atttttaatc   | agtaaaaacta | aataccacaa   | 63840 |
| caacgggtta  | ttggtcttaa | atttcgagac  | atctttcttt   | ataaaatagt  | cacttggaat   | 63900 |
| taatccagca  | agccacctca | tactaacgag  | actaaataaa   | aaaataaata  | aagatatattga | 63960 |
| gctcagccaa  | agtataagaa | ctttataaga  | ttcaaaaaaaa  | attatgatgg  | catttaggggt  | 64020 |
| attaaattcc  | ataattaccc | acctttttat  | ttaataatga   | tttagtatag  | caatgtctaa   | 64080 |
| aaataaaattt | aaacatttaa | taaatatata  | ataattttgc   | caaaaaaagc  | cttaaatgat   | 64140 |
| atttgcactg  | aaaccctcag | tgataataaa  | gccgaaaaatg  | ttctatcact  | tgatatcaaa   | 64200 |
| ggtatttctt  | cttttgctga | caacatcatt  | attgcaactg   | caaattcgaa  | taggcattgca  | 64260 |
| aagtcctttat | ccgaaaaagt | agttgaggag  | ataaaagcta   | ataaaattag  | tatcatgggt   | 64320 |
| gtagagggca  | agacagaatc | aggttggata  | ttagtagact   | gtggtgaggt  | tgtagtaaat   | 64380 |
| attatgaaga  | atgacataag | agagtttttat | gatttagaag   | gtctttgggg  | tgaaaaacacc  | 64440 |
| ctcatcgatt  | cttcgaaata | aatgctatta  | aatatcataa   | gtgttggaag  | ctcaccttca   | 64500 |
| tcttggggagt | taacaggaat | agaatattac  | acaaagcaaa   | tccttaaga   | agtaagtcta   | 64560 |
| aattttatta  | acgtaaaagg | gcagcaacac  | ccaaaaagat   | ctacagaaga  | ggtcttaaaa   | 64620 |

cttgaatcta agttaatctc gtccaaaata gattctaata gataatattgt ttgctgggac 64680  
 tcaagtggcg agtcgttgaa taatttttgaa tttagtaaat tttttgaaaa atccatgctg 64740  
 gaaaatatga agctttactt tgtaattgggt gggtcatttg gcatacctcg agatattctt 64800  
 gataaatcta ataagataat ctcgatctca agtctcaccc tacctcacag gcttttcaag 64860  
 attgttctta tagagcaaat ctataggtca ttttcaattc tcaaaaatct tccctaccac 64920  
 aaatgattga tctgaatgag aggtttgttg aaaaaagaag ctttttcaat agattgttgc 64980  
 ttatatattt tttctttggg gccctgtttt tattcttgct tttcaagacc tattctttac 65040  
 aggtttctag ctattctgac tatgaactag cagcactaaa gaataaaaca aaagaagtgc 65100  
 tagttcagcc tgttcgagga gtcattctatg acaggaatgg aaatatctta gtttaataatg 65160  
 tccctacata tgacctaaata atccaaccat caaaaattaa aaacttagat gatttcatta 65220  
 ttaatatattc taaaaataata gagctatcag attcagaaat tgaaaacatt attaaaaatt 65280  
 ttaaaaaggag tgcaacttat aatagagagc taacaattaa aaaagatctc tcgaaagaag 65340  
 aaatagcaaa atttgaagta agaagctacc aatttcagaa tgcattata gatgttaggt 65400  
 atagtagaga aaataaaatat ccttacttat tttcacacgc actcggttat gtaggtggtg 65460  
 taagtaatga taaggtttta tcaattttta aaaatcaaga tttgaagcaa tctgaaacaa 65520  
 ctttttaagta ttcaggtggt tttattgctg gaaaaacagg attggaaaaat atttatgatt 65580  
 cagctttaag aggttccttt ggaaaaaagt tatttgaagt tgatgcaagg ggaaggcctt 65640



|  |       |
|--|-------|
| taaaagagct aagttttgaa aaacctatta atggaaaaag ctttttact catttagatt   | 65700 |
| taaattcaca aaaaaagcc ttgaaacaaa tgaataatag aagggtgct gttgttgccct   | 65760 |
| tagagcttaa atctggttct atagttacct acctaagcac tccaagcttt aatgtaaatg  | 65820 |
| gtctttctaa tggcatgtcg tcggtagagt tcgaaaaact gattaatgat gtagacaagc  | 65880 |
| cattttttga tagggccggt caaggtcgggt actctccagc ttctacaata aagccagcaa | 65940 |
| tagcattgtt tggattataa gaaaaaatag tagactggaa ctttacctt aaagatcctg   | 66000 |
| gattttttgt attaccagag gatcagagga ttatataggg atggaagaaa ggaggtcatg  | 66060 |
| gaacaaatga tatgaagggt gcaatcatag aaagtctctaa tacttttttc ttttctcttg | 66120 |
| cttataagcc tgatattaat aatttaata gccatctttc tgagtttggt ttbggcagga   | 66180 |
| atgttttgtaa agattgtttt aatccagatt ctggattatt gcctacgcct gaatggaaaa | 66240 |
| tgaataatct taattttggc tggtttaaag gagataccgt taatttaggt gttggtcagg  | 66300 |
| gctatatgag tgccactcca gttcaattag catattactc tgcattttctg gcaaaaaaag | 66360 |
| gaaatcttca agagctatca ttgtgcgaga gtgacagcct aagcaatact gcttttataa  | 66420 |
| ataatttaag catagataac tcagactggg atcaaaattca ctcaagcatg attggagtta | 66480 |
| ttgaagatcc gagaggcact gctaaaagat taaaaccttt aaaatcataat gttgttgctg | 66540 |
| caaaatctgg aacggttgag cttgtcagca cacaacaaaa ggaagattac aaactagtaa  | 66600 |
| ggcaaaaatat tggtaacaga gaccatgcaa tcattgtggc ttttgggtcca atgcctgac | 66660 |
| cagagtatgc agtaagtgtt gttattgaaa atggtgaaa cggcggttct gttgcaggtc   | 66720 |

|             |            |            |               |               |            |       |
|-------------|------------|------------|---------------|---------------|------------|-------|
| ctgttgctat  | tgctgtttta | aatgagctta | tcaataaatg    | aagaaaaaat    | tagacttta  | 66780 |
| aaacttttagc | atttattttg | atcaatat   | tttattg       | ataaccctgt    | tgtcggtcat | 66840 |
| gggtttattt  | ttttatata  | gcgcataca  | agaagata      | tc agcactgttg | ctaagcaagc | 66900 |
| tgtattcgtt  | ggtttttggt | tgctgtta   | at gtttgtagtt | agccaacctg    | accctgattt | 66960 |
| ttataataca  | ttttctgggt | tatttttttg | ggggggastt    | gtattgattt    | ttctaactat | 67020 |
| gatttttgg   | aaagaaaata | atggagccaa | aagatggctt    | gatttaggat    | tttttaccct | 67080 |
| gcaatcctct  | gagattatta | aaatatcatt | gccaaattttt   | ttatcatcat    | atttatataa | 67140 |
| taagccactt  | ccaataagca | ctaggcatac | ttttattaca    | ttgatattaa    | taggttttat | 67200 |
| atttgcactt  | gtagctagac | aacctgactt | gggaaccagt    | ttagtagtat    | ttatgtcagg | 67260 |
| gggttatgta  | ttatttttag | caggattaag | ccgccgttta    | ttggatctgc    | aatagcctta | 67320 |
| tttttattat  | ccttacctbt | tttatggaat | aactttctag    | agccatttca    | acaacagaga | 67380 |
| gtttttaactt | atttgaacc  | agacgcagac | ccttatggta    | ctgcatggaa    | tataactcaa | 67440 |
| tcaaaaaatag | caattggatc | aggcggaatt | aatggaaaaag   | gttatcagga    | gggctcccaa | 67500 |
| gcccatcttg  | atdttttacc | agagacagaa | acagattttta   | tttttgctgt    | tattgctgaa | 67560 |
| gagtttggtt  | ttattggagt | ctgtattttg | ttatcagtat    | tttcttttat    | atkactcaga | 67620 |
| tgtttatatt  | tagcatttaa | tgcaagagat | agatttttgc    | ggttaactat    | aggaggccta | 67680 |
| agtttagttt  | ttgcctctac | attatttatt | aatttagcaa    | tggttgttg     | tgtagttcct | 67740 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| gtagttggta  | tgcctcttcc  | atttatcagc  | aaagggcggt  | catctttgct  | atcctkttat  | 67800 |
| atagcttttg  | ggattataat  | ttctatggca  | acacataaaa  | aattaatgca  | aagatgaaaa  | 67860 |
| aaattatat   | tataacttta  | atattcacta  | tttcaattac  | tgcggattat  | tcgaatcatg  | 67920 |
| aagatagcca  | aatggtaata  | aatgaactcg  | tcacaaaaa   | cggttttgag  | gaatcttatg  | 67980 |
| ttactgcaat  | cttaaaaaat  | gcaaaaaaagc | gtgatgagat  | gcttaaatct  | gttgctaatac | 68040 |
| ctgctgaaa   | aacaaaaaaca | tgggatgaat  | atagggtat   | ctttataaaa  | acaaaaaagag | 68100 |
| tttcagaagg  | caaaaaattt  | ataaaaaaaa  | atattaatgc  | tttagagagg  | gctgagaaaag | 68160 |
| aatttggagt  | tcctaaagaa  | ataattactg  | ctatttttagg | cgtagaaaact | aattatggca  | 68220 |
| gcaataaaagg | tggatacaga  | gttttagata  | gtcttactac  | tttaggtttt  | gatgacccgc  | 68280 |
| gaagatctaa  | ttttttttaga | agagaactta  | tagagtctct  | ccttttaaca  | agagaaaaata | 68340 |
| atttagatat  | caaaaacaaca | aaaggctcat  | atgcgggagc  | catgggatat  | gcacaattta  | 68400 |
| tttcgtcaag  | ctaccgagcc  | tatgctgtag  | attttgatga  | agatgggttac | gttgattttat | 68460 |
| ttaattctgt  | tgatgatgca  | attgggagta  | ttgcaaaatta | tctttatgtt  | catggatgga  | 68520 |
| agagagaagg  | aaagatcgta  | acaaaaactt  | atccaaaaca  | tgttagaaaa  | ttttataaac  | 68580 |
| ctcatgagtc  | tctaacaagg  | ttcatacctt  | taatcttta   | tgaagatgga  | aaagatcttt  | 68640 |
| tttttattgg  | tgatgataac  | tttagggcta  | ttgctaagta  | taatattagc  | gatgtctatg  | 68700 |
| caatggcggt  | ttattactta  | tcagaagagt  | taaaaaaatg  | aaaaaattat  | tattcacatt  | 68760 |
| actatcaact  | tcaatatatta | ttcaagcaca  | gagttttgtt  | ccagattctc  | ctgaggttaga | 68820 |

|             |            |             |             |             |             |       |
|-------------|------------|-------------|-------------|-------------|-------------|-------|
| tctgaagagc  | tatacctaa  | ttgagccaaa  | caccaatact  | gttattgcag  | aatttaattc  | 68880 |
| agatttggaa  | atagaaccag | ctagcatgac  | taaaatttatg | actagctatg  | ttgtggctga  | 68940 |
| tcagattgca  | aatggtttaa | tatctcttga  | tgacgaagta  | ctaattagtg  | aaaaagcatg  | 69000 |
| gagaatgcaa  | gggtctaaga | cctttataga  | agctggaaaa  | aaagttacgg  | tatctgatct  | 69060 |
| tttaaaaggt  | attatgattc | agtcaggcaa  | tgatgcctct  | gttgcaattg  | cagaatatgc  | 69120 |
| aggcggcact  | gaaagaggct | tcgttgactt  | aatgaactct  | tatgctgcct  | ctttggagat  | 69180 |
| gaataatacg  | atctttcaga | attcaacagg  | gcttccagat  | gaaaatcatt  | tttcatcagc  | 69240 |
| aaaagattta  | gccaatctga | cagctaacta  | cattaataaa  | ttccagaaag  | aatatgcttt  | 69300 |
| atataagcag  | aagcaattta | cttttaacaa  | tattaagcaa  | ttaaatagaa  | ataagccttt  | 69360 |
| atggagagac  | gactcagcag | atgggtgttaa | aactgggtcat | acagaagctg  | ctggatactg  | 69420 |
| tttagttggc  | tcagcaaaaa | ggggcggtat  | gaggcttatt  | acagttgttg  | caggaagtaa  | 69480 |
| gtcagataat  | gatcgttttt | tatcttccca  | aagattactt  | gagtatggat  | ttagattttt  | 69540 |
| tacaacacag  | aagatgttaa | gtgcagaaaa  | agaatatcag  | aatattacag  | tttgggggtg  | 69600 |
| acaagaaaaag | atacttggcg | tgggagttct  | agaggatata  | tctattactc  | ttcctagaac  | 69660 |
| aagtttcaaa  | aatgttgaaa | ctgtttataa  | agttaataac  | aatatccagg  | ctcctatcgt  | 69720 |
| agttgggtcaa | aaagtcggga | ctctagaaat  | aattagtaat  | gatgagattg  | ttctagttac  | 69780 |
| ggatttggtt  | gctttaaaaa | atattgaagc  | taaaggtttt  | tttgggaagaa | tatgggtcaaa | 69840 |

gttcgttctt tggatatatta gcttatttgg tctaacagat gaaaacacca cttaaaggcg 69900  
 tttttaatgg ggcttttgat actgtagata acataaaaaat atcacctttt tctcgagctt 69960  
 atactttttc tgatagcgta tatgaagtag ttcttttctt taattcaact gcaattgctt 70020  
 ttgatgatca tataaaaaa ttagaatttt ctgctagtca actagcgatg gccgtagatt 70080  
 taaaagaagt tataatttgaa attaatctt taataacatc atctgaattt agtaattggct 70140  
 atgtctatta tcaagttact cgaggcgttg atccattaag gtctcatatt catgagccaa 70200  
 atttaaaaat agaaaaccttt ggctatgcaa aggcgcactt attccaatgg aaaccattaa 70260  
 gagtatcagt atgtgatgat attaggtggg ggaggtgtga tattaaatca acatcattgc 70320  
 ttggtaatgt tatgaacatg aatgcagcaa aactagataa ttgtgatgaa gttattatgc 70380  
 acaaagataa tttgttgacc gaggtggcg catctaattc tttttttgta aaaaatgatt 70440  
 caatctgcac ccagctcta aatggaaata tacttcccg cataaccaga gcattactaa 70500  
 ttaatgaatt aaaaaattat agtatagagg tcatcgaaaga taacttcagg ctagaagatt 70560  
 tatcaaatgc ttcttgcgcg tggctcacia gcgcaacgaa agggctggcc cccatttctg 70620  
 aaattagtaa cctagaatct cacttggatt tataccatcc tcttttcaaa aagagtgaag 70680  
 aaatttttaa taagaagtcc cttagttaga agtcagttta ataactatat tatcaagctc 70740  
 gtaccaaaaa tctttatcac taagaccttt atttgcaaga tcaagctcaa atacattttt 70800  
 ttggagctgt attaaagttt gtagtttatg atttttaata aaatgcataat attcagaagt 70860  
 tttatttttc caaaccaccag atttaattag actatctaca ggatttttac ttgtcttagc 70920

|             |            |             |            |             |             |       |
|-------------|------------|-------------|------------|-------------|-------------|-------|
| attggctgca  | ctgttaataa | ttttaccaac  | tatccaaact | agaagagggtg | cataatgatc  | 70980 |
| ttccgctgat  | tttattgaat | ggataatttt  | taaagcttgc | ttggtattat  | ttgaaataat  | 71040 |
| tttatcttca  | agctcaaatg | gcataaatc   | tgcagactcc | acagaagggt  | ttctttcttg  | 71100 |
| gccatcacca  | tctttataag | tcagtcttag  | aagctttact | tcattctgtt  | gggcaactaa  | 71160 |
| gtttccagag  | ttcatttcag | atatattttg  | cactaaagag | ggcctatctt  | tttcagaaat  | 71220 |
| aaaagatagt  | tgatgcttaa | gccaaatctt  | ttcttcaaaa | gatttaagtt  | ttccacaatc  | 71280 |
| aatgatgagg  | gctatctcat | ccatctgttt  | taccacttta | gttgtgaaat  | taagcttttc  | 71340 |
| attgcttgaa  | ttaataatta | ttgcaatatt  | atccatatgg | ttaatatttt  | caatttgaga  | 71400 |
| tattttaatt  | atttggtcag | gtattttttcc | ttgatcatga | ttaatctcaa  | ctattatttt  | 71460 |
| tgagccaaat  | aaagatccag | aagcattttc  | aataatagtc | tgatctattt  | tatcaaaacc  | 71520 |
| atctttatta  | ataattgttt | tttctttgaa  | ccctttattt | gataaatgtt  | tgagcaaaaag | 71580 |
| atcttttagag | ttatttttta | agacaaacctc | agatccaaaa | ataaagaaga  | tatttttgtga | 71640 |
| ctgatctaaa  | tatttttgag | ttgtcattgc  | ctcacatttc | aaaagcatgc  | acctcaaaaa  | 71700 |
| ttatttgatc  | aattaattcc | ttctgcattt  | cagactcaag | agattttaatc | atttgatctt  | 71760 |
| gtgcaaatgg  | attaagttca | ttcgatttat  | atctcttgta | gctaacaagc  | ttttactta   | 71820 |
| tttgtttatt  | tccaatcatt | tttatatgta  | tttctaaatt | aatttctcct  | tcaagagctc  | 71880 |
| ttagagaaga  | gcctccatag | atatcatatt  | tatttaaaat | atagtcttgt  | attagaatct  | 71940 |

|            |             |             |             |             |             |       |
|------------|-------------|-------------|-------------|-------------|-------------|-------|
| gattttgacg | attctcagct  | tctttttctg  | tattgaagag  | gattgctaac  | ttttgctcaa  | 72000 |
| aagaattagg | gacactacta  | tcaaaccgaa  | aattaaatag  | atctttttta  | tctagatcta  | 72060 |
| tatttagttg | attaaaattgg | cagccagtta  | aagatatcaa  | caagcaagtg  | tataaaaaa   | 72120 |
| aaggctctag | tttcataaacc | tcatagtatc  | tatctgaaga  | atataatatca | aagattaatt  | 72180 |
| ttaaataaca | aaatttatta  | tttttttcttt | aatataaata  | gttttttttaa | tagccgagtt  | 72240 |
| gtttattgtt | gatgccacat  | tgtcgatagc  | cagtgctaaa  | gcttctatat  | cttttttgctc | 72300 |
| aaggtttttg | tctatcatta  | ccttacctct  | cacctttcca  | tttacttgaa  | ctattaaatt  | 72360 |
| aaattcagaa | acttctaata  | actcctctct  | aagaacaggc  | caagaagatt  | caattttcttc | 72420 |
| ttgtgcaaag | tcaaagtaaa  | aattattcca  | taaatgttgt  | gagatatgag  | gcgcaatagg  | 72480 |
| attgagagtt | ttaagaataa  | ttattattgc  | ttcatttaaga | caatattgat  | ttgagatcga  | 72540 |
| cgcattacct | tctttaaaag  | aatcagggtat | aaagttaact  | agctccatta  | tcgaagcaat  | 72600 |
| ggctgtatta | aatgaatatc  | ttgtttcaaa  | gtcatttgta  | accttttttaa | gagtattatg  | 72660 |
| cgattttctt | cggagctcta  | cttcttgttt  | cgaaggatcc  | ttgggctcct  | ctaagtcaat  | 72720 |
| aaattttcta | ttgctgacta  | agttccatat  | ttttttcata  | aatcttgagg  | caccttctac  | 72780 |
| agatgactct | gaccatttcta | gacttttgctc | agggggagct  | gtaaacatca  | tgtaaaagcct | 72840 |
| tacagtgtca | gcccataact  | tttctatata  | agattgagga  | tctacagtat  | ttccttttga  | 72900 |
| tttggacatt | ttagccccc   | ctttaagaac  | catgccttgt  | gtaagaagct  | ttttaaaagg  | 72960 |
| ctcattgcct | tcaactaaat  | ccatatctct  | aagcgcctta  | tgaaaaaaatc | ttgaatatag  | 73020 |

|   |       |
|---|-------|
| taagtgcagt atcgcatgct caatacctcc tatatataaa tcaacaggca gccaatattt   | 73080 |
| tgaatttttta tcaaacattt catctgcatt gtcagatgaa gtaaatcttg cgtaataacca | 73140 |
| tgatgagtc ataaatgtat caaaagtatc tgtctctctt ttagtcgat cagaaaatatt    | 73200 |
| ataaaaaatct tcattttgac ttaagggaat agtgccgag ttctttttta gctctggaag   | 73260 |
| cactataggc atatcctttt catcaataac tcttggttca ccattttcat aaacaaccgg   | 73320 |
| aatagggcaa cccaatatc ttgtctgct cacaccccag tctcttaatc tgaactggat     | 73380 |
| taactgctca ccaaggtttg cgtctgctaa atctttaata atttcaagag atgcctcatc   | 73440 |
| agagtccatg ccatacat tcatcagagt tattaatttg cctttttgta caactggaag     | 73500 |
| ttcattatta ttatcaaac taataacttg aggaatctct agattatat tagatgcaaa     | 73560 |
| ttcgaaatct ctttgatcat gagcaggaac acccatcact actccagttc cataatcaag   | 73620 |
| aagtacaaaa ttccaatcc atacgggtat cttcttcttt gtaattggat gaattacatg    | 73680 |
| cattccacta ttaattccaa gttttttcagc tttagccata tcagcttcag cagctttcac  | 73740 |
| ttctttgcat ttgtttagaa agtcttttat cgattcata ttttttgaca accctattga    | 73800 |
| tatagagtga ttgggtgata ttgctaataaaa agaaacacca aaaattgtat caggtcgtgt | 73860 |
| tgaaaaatacc tttaaagaat catcggaatc ctcaatcatg tatttaattt ctgctccctg  | 73920 |
| agattttcca atccaattcc ttgcatgtgt tttaacattt tcaggccaat ccacctcatc   | 73980 |
| caaagatgtt aaaagttcct cagcgtagtc tgtaattttt ataaaccatt gatcaatttc   | 74040 |



|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| ttttatttca  | acttgcgccc  | cagacctcca  | tcctttttcca | tctataaactt | gctcatttgc  | 74100 |
| taaaactgtc  | tcatcaactg  | gatcccagtt  | cactaatgat  | tttttcctat  | aaaccaaacc  | 74160 |
| tttatcataa  | aacttttttaa | atattaaactg | ctcccactta  | tagtattccg  | gttcacatgt  | 74220 |
| tctaagctct  | tttgaccaat  | catacccaag  | gcccaagat   | ataagttgct  | gtttcatatg  | 74280 |
| ttcaaatattt | tgattcgtcc  | aatctttttgg | gctaaacttta | ttagcaatcg  | ctgcgtttttc | 74340 |
| agcaggaagc  | ccaaatgcat  | cccagcccat  | aggctgaaaa  | acattaaagt  | tattcattct  | 74400 |
| tttatatcta  | gatatgacat  | caccgacgt   | atagttttctt | acatggccca  | tatgtaattt  | 74460 |
| acctgatgga  | tatggaaaaca | ttgataggca  | ataaaaatttt | tctctaccat  | caggattagc  | 74520 |
| tttaaatttg  | tcttctttta  | accattcatc  | ttgaatgggt  | ttctctattt  | gacttgggtt  | 74580 |
| atatttcagga | ttcattttttt | taggaactca  | tttttctaagt | aattaatagt  | atgttttat   | 74640 |
| tctttaaaagg | tatcatcatg  | cagaatcctt  | tgatgaagag  | ggtgatttgt  | tttaatgccc  | 74700 |
| tctataaaaa  | attcatctaa  | agcactcagc  | attctcttaa  | tagccgagct  | tctagaattt  | 74760 |
| gcctgagtaa  | tgattttttgc | tagaagtga   | tcatagttag  | ggggaactct  | atatacctcca | 74820 |
| taaaatatgtg | agtcatatct  | tatgccaaaa  | ccgccagggtg | tatgcatttt  | tgtaattgtt  | 74880 |
| cctgggggatg | gttgaaaaatt | atcaggatct  | tctgcattaa  | ttctgcattc  | tagtgagtg   | 74940 |
| ccatgaaaaat | taatatcatc  | ttgatttagc  | tcaattggca  | tttcaagagc  | aattcttagt  | 75000 |
| tgtgcctttta | ctaaatcaaa  | gccagttatc  | atttctgtta  | ccggatgttc  | aacctgaatt  | 75060 |
| cttgtattca  | tttctataaa  | atagaattga  | ttatcttctgt | ataaaaaattc | aatagtgcc   | 75120 |

|  |       |
|--|-------|
| accccttcgt aatttatttc ttcacataaa ttgacacaaag ctattagtgt tttattaaga   | 75180 |
| gcttcttgat ttatattaag tgctggagct tcctcaatta ttttttgatg tctcctctgc    | 75240 |
| atactgcaat cccttgtaac taaatgtatc gcctttccctt taccatcacc aactatttga   | 75300 |
| acttcaatat gtctagggtt cccaataaat ttttctagat aaatggtttc attgccgaat    | 75360 |
| ccatttttag cttccctgcat tgttatttct gcatgtccaa taagatcttc ctctctttca   | 75420 |
| acaactctca ttcctcttcc accgccacca gccgttgcct taatcataat tggatagccg    | 75480 |
| atatctctag caattttttt aaactcatca ttatcagatg gaatctcatc tttataaacct   | 75540 |
| ggaacgattt gaattccaga tttttcagcc agtgtttttg ctgttatttt gtcgcccat     | 75600 |
| ttttggattg tttctgaggt tggcccaata aatttaaadc cacttttttc acacatttct    | 75660 |
| gcaaaaattat gatcttctgc aaggaaacca tatccaggat atatagcatt agcacttgta   | 75720 |
| agttctgcag ctgagagaaat ggcagggtata tttaggtagc tttagagttgg agatgcaggc | 75780 |
| ccaatacaaa cagtttcatc agagaatctt aaatgtttta gatctttatc agcttctgag    | 75840 |
| taaacagata ctgtcttgat accaagctct ttacaagctc ttatagctcg aagagcaatc    | 75900 |
| tcaccacgat tggcaataag aactttataa ctcattttaa ttgacagtta taattttttg    | 75960 |
| tccaaaattca accggacttc catcttccac gtcaataactt gaaatcttgc catcaaatc   | 76020 |
| tgattttaatt tcattcatca ttttcatggc ttcaactata cagagcacat cacctacttt   | 76080 |
| tatatgtctt cctactttta caaatggatc tttttcggga cttggggttc tataaaaaagt   | 76140 |

|             |             |             |            |             |             |       |
|-------------|-------------|-------------|------------|-------------|-------------|-------|
| tcctactatt  | ggagaggtaa  | caacatcccc  | tatgacagtt | tgcttagcct  | catccaattg  | 76200 |
| atgagttggt  | ggtgagatag  | ctggctcatt  | tttgacaatg | attggttttg  | aaacaaaattc | 76260 |
| ttgattagag  | cggttatctc  | ttgaaaattct | tactgactca | tctccttgac  | ttactttctat | 76320 |
| ttcttttaaa  | tcagactctt  | gaagcatttc  | tataagtttt | tttattttcc  | taatatccat  | 76380 |
| ttattttacc  | ctgtatttta  | atatgccctt  | aagcatcgcc | tcttcataac  | ccttagcacc  | 76440 |
| aataccagtt  | atcactttct  | cagctatatc  | agatagatat | gagacatgtc  | taaattcctc  | 76500 |
| tcttgatat   | atgtctgaaa  | ggtgaacctc  | atagaatggt | atgttgacac  | caagaaatgc  | 76560 |
| atcccttata  | gctatgctcg  | tatgagtata  | tgcagcagga | tttattatta  | caaagtctat  | 76620 |
| tttttcttca  | atagtgtctat | gtattgattc  | tattatttca | tgctccgcat  | tactttgaaa  | 76680 |
| agataattaaa | ttgcatttgat | tttgattttgc | taagactaat | aaatctttgtt | gaatatcttc  | 76740 |
| aagtgatttg  | ctgccatata  | cttccttttc  | tctagtccg  | agaagattga  | ggttggggcc  | 76800 |
| atttataaag  | aatatattca  | tataatagat  | tttaacagaa | atttacagat  | tttttaggat  | 76860 |
| tttaaaagata | tttttataag  | agttgatttt  | tttgcatggg | atagcaaaaa  | cccgcttcag  | 76920 |
| agcatccctg  | atagtatat   | aaaacctccg  | ataagttcaa | tgaagaattt  | atttcaatgc  | 76980 |
| ttataaataaa | ttcatctcta  | agaatctctag | tctcgccaaa | aaattcatcc  | ctatatagat  | 77040 |
| ctttacttga  | ctcaagcgtc  | ttaaatggga  | gcactttatc | tttaaatgtg  | agctctatac  | 77100 |
| tatccaagta  | catataataa  | ccatctctcta | tttgccaagt | tatatatgct  | gcttggtttt  | 77160 |
| caacattggt  | tgtaaaagca  | aatacttcat  | tggtttcttg | gacctatatt  | gagttttcaa  | 77220 |

|   |       |
|---|-------|
| aaagattgc tgaattcaat tcacctgcaa tcacattgga tagtagtata attattaaac    | 77280 |
| ttttgatcat gaagacttaa gaataaaccc ctaagtactt aaaataaaga ttttattaaa   | 77340 |
| taacacggta gttaaaaaatt tatgttccag actttcgcac taaaaatttt cttttatatg  | 77400 |
| cctatttggc ttttaaaaat aattttttat agaaaaagaa cagtaataag agggcatcaa   | 77460 |
| tttgatgctc aatctgctgc attattgtct ttgctttccaa agaaaagattt atctgaatta | 77520 |
| ttagatggtg aaattgcaaa agctagaatt actcttgaag aagcaagaat tcaaaaataaa  | 77580 |
| gttttctttaa caccgtctat acaagtcaga aaagtagatc atattttacc aaagcatgat  | 77640 |
| ttaattctca gagaatacaa gccgcaccaa gaggatttaa aaaaagtatt tctctatttt   | 77700 |
| catggaggag gttatgtcct taactcagtc aatacacatg atgatatggt ttcatatatg   | 77760 |
| tcagataaaa taggagtaaa gttttattct ctagactata gactatcgcc tgagagcaaa   | 77820 |
| tatcctgatt ctcttgatga tgctctagat gctttttctt ggctcatcgg ccagggatat   | 77880 |
| ggaccagggt atatttccagt ttgtggagac agcgcaggcg ctcatctagc tgcttctttg  | 77940 |
| tgccattacc tagctgaaca aaataaaagat ttgcctagta gccagctact aatatatcca  | 78000 |
| atgtgtgatc cgtcctgttc atctgagtct tataatttat tatcttcagg atatcttcta   | 78060 |
| actaaaaaaaa ctatgatttg gttttgggat aagcttagaa ctcccgagt aaacaataac   | 78120 |
| gattcggcct ttaatctctt aaaatttaat ttgaaaaaaaa ctttaccgcc aacaattatc  | 78180 |
| gttacagggt gctttgatcc tctatgcgat gatggagaaa aatatgcata tttattacat   | 78240 |

|            |             |               |            |             |             |       |
|------------|-------------|---------------|------------|-------------|-------------|-------|
| aaaggtaaac | ataatgtgaa  | acaattacat    | tatccaacaa | tgtttcatgg  | gtttgcatca  | 78300 |
| atgactaaat | tgaagcagc   | gcagatagcc    | gttgaagatt | ttttaaaga   | atataagaaa  | 78360 |
| atactatgag | taaaatttta  | gaagtaagt     | acttaagcat | caatttttca  | acaagggatg  | 78420 |
| gattgttta  | tgcagtggat  | aatataagtt    | ttgatataga | aaaaaatcaa  | accttggcct  | 78480 |
| tggttggtga | gtcaggttct  | ggtaagtcgg    | taactgctat | gtcaattctt  | cagctccttc  | 78540 |
| aaaaaccaca | agcatcatat  | tccaaagagt    | cttctattaa | gtttaatggc  | gatgagataa  | 78600 |
| taaatgccaa | gtatgaaaag  | ttactttcct    | tgagaggaaa | tattatatct  | atgatatttc  | 78660 |
| aagagccgat | gacctcacta  | aacctttatc    | acagagtagg | taatcagata  | actgaatcaa  | 78720 |
| tactacttca | ctcaaaaagc  | tcaaaaaaag    | atgcaataga | tgaagcaaaa  | aaattaatgg  | 78780 |
| cacttggtga | gattgatgat  | gttgaaagac    | ggttctatgc | ataccctcat  | gagcttttctg | 78840 |
| gagggcagcg | acaaagagtt  | atgattgcta    | tggcccttgt | taataaaacct | gagctattga  | 78900 |
| ttgctgacga | gccaaacaacc | gctcttgatg    | taactatcca | agcccagata  | ttagatctca  | 78960 |
| tgtctaagct | taaaaaatgaa | ctaggcatgt    | caatactttt | tattactcat  | gatctaggcc  | 79020 |
| tcggtcaaga | attttctgac  | aatgttttgtg   | ttatgaagaa | tggcaagata  | gttgaacaag  | 79080 |
| gaaatactgt | tgagggtat   | ttt aacaaccct | ctcacgaata | tacaaaaaaa  | cttttagatg  | 79140 |
| cagaacctca | gcccaaatta  | gataaccctg    | taagtgatga | gccataaatt  | gagatcaatg  | 79200 |
| acttaaatgt | ttactactca  | ataccttcaa    | ctaatttttt | taaaaagaat  | acttttcatg  | 79260 |
| ctgttaaaaa | tacttccttt  | agtatttata    | aaaatactac | aattggcctg  | gttgggggaat | 79320 |

|             |             |            |             |             |             |       |
|-------------|-------------|------------|-------------|-------------|-------------|-------|
| ctggatctgg  | aaagtcaact  | ttgggtaagg | ctatagcaaa  | ctagtttct   | tataagggga  | 79380 |
| atattaaatt  | tgagggaaga  | gatatcaact | caaattccca  | aaaagaaaaat | aaagaattga  | 79440 |
| aaaaaaaaat  | gt          | tttcaagatc | cttatgggtc  | attatcacca  | cgaatgacag  | 79500 |
| tgggggagat  | agttggtgaa  | ggtttaggtg | ttcactttta  | gcttacaaaa  | aaagaaaagag | 79560 |
| acgaaaggat  | agataagggt  | ctgtcagatg | tcggtatcga  | aatagtagct  | aagaataaat  | 79620 |
| atccgcatga  | gttttctgga  | ggccaaaggc | agagaattgc  | aattgctaga  | tctttaataa  | 79680 |
| tgaatcctgc  | tttttatgatt | cttgatgagc | caacatcagc  | attagatagg  | tctatttcaaa | 79740 |
| ttcagggtaat | cgattttatg  | aaagagatac | aaaatgaata  | tgggcttact  | tatcttttta  | 79800 |
| taagtcatga  | tttaaagggt  | attagatcga | tgtcagactt  | tatttttgtt  | atgaaaaaatg | 79860 |
| gagaaatcgt  | agagtcagga  | ccttctcaca | aggtctttga  | aagcccagag  | caagactata  | 79920 |
| ctaaaaaatt  | actatcagct  | gctttaaagt | atgcatctga  | ataattaaat  | atatymcmtm  | 79980 |
| tgggcaaatag | aaartattca  | aaagagctcg | ttgacgggtcc | taatcaagct  | gcttcttagat | 80040 |
| caatgcttag  | aggagtaggt  | ttcacatctg | aggmtttcac  | aaaaccattt  | gttgggattg  | 80100 |
| cttccacagg  | agcaaaaagta | accccatgca | atatgcacat  | aatgcactt   | tcagagatcg  | 80160 |
| ttgagaaaatc | agttgatagt  | tcaggaggaa | agggtgttct  | ttttaatact  | attactgttt  | 80220 |
| ccgatggaat  | ttctatgggt  | acacagggta | tgaatatattc | tcttgtttct  | cgagaggtaa  | 80280 |
| ttgcagattc  | aatagaaaact | gttgtgggat | gtcttggtta  | tgatggagtt  | attgctgtcg  | 80340 |

|             |             |             |            |             |            |       |
|-------------|-------------|-------------|------------|-------------|------------|-------|
| gtggttgtga  | taaaaatatg  | cctggatgca  | ttattggaat | ggcaagatta  | aatagaccat | 80400 |
| caatatattat | atatggtggt  | tctatcaaac  | ctagtaaaga | aaataccgac  | tatgttactg | 80460 |
| tttgtgagaa  | aactggagag  | tactcaaaaag | gcgatcttaa | agaatctgaa  | taaattcatg | 80520 |
| tagaaaaaat  | ttccgtaaaa  | gggcctggat  | cttgtggggg | aatgtatacg  | gcaaatacta | 80580 |
| tggcttctgc  | gattgaagct  | ctaggcatga  | gtcttccttg | aagcagcagt  | caagatgcaa | 80640 |
| tttcacacga  | caaagaagat  | gattgtttta  | aggctggcga | agcgataatg  | aatttattag | 80700 |
| aaaaagatat  | taagccttca  | gatattatga  | ctaagaatgc | ttttgagaat  | gctataacaa | 80760 |
| tggtaatgct  | tctaggaggt  | tcaactaatg  | cagttcttgc | tttatgggc   | atggcgcat  | 80820 |
| caataggggt  | tgatttagag  | ctagatgact  | ttacaagaat | aggaaaaaaa  | acacccgtta | 80880 |
| tggcagatct  | taagcccttt  | ggttctcatt  | atatgtctga | actcaatgct  | aatggcggta | 80940 |
| ttcagccact  | aatgaaaaact | ttgcttgaga  | aggattact  | acacggcaat  | tgtcttaccg | 81000 |
| ttaccgggtca | gacgcttgct  | gaaaatcttt  | ctggaataaa | accttacgag  | cctgatcaag | 81060 |
| agataaattaa | atcatttgat  | aatccaatta  | aatcaaatag | ccatcttaga  | attctgtatg | 81120 |
| gcaacttagc  | gaaagatggt  | gcagttgcaa  | aaattacggg | taaaagaagg  | acttcctttg | 81180 |
| aaggaaagtgc | tcgtgtat    | gattcagaag  | aagaaggggt | taaaagcaatc | ctatctaaat | 81240 |
| ctataaaaagc | tggagatgtt  | gttgtaatta  | gatatgaagg | gccaaaaagg  | ggtccaggca | 81300 |
| tgagagaaat  | gctaaaaacca | acatctgcc   | taatgggtca | aggtcttggc  | gacaggtag  | 81360 |
| cttttataac  | agacggtcgt  | ttttcaggag  | gcactcatgg | attgttgtt   | gggcacatta | 81420 |

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| ctccagaggc  | tgcagatggc  | ggcttaattg  | caataattaa  | agatggcgac  | tctatatata  | 81480 |
| ttgatgcaga  | ttctgataag  | ttaattctta  | atattttctga | ggatgaaatt  | tcaaatagac  | 81540 |
| taagtagatg  | ggtaaaccaca | aaaacgcctc  | ccaaaaaagg  | agtcttagca  | aaatttgcaa  | 81600 |
| aaagtgttaa  | atcagctagt  | cttggagcgg  | taacagatta  | aatatgtatt  | ttaaaaagaaa | 81660 |
| atttccaaat  | agtagactaa  | gaagaatgcg  | tctcaattca  | aacctcagag  | acttgcttgc  | 81720 |
| tgaagttagg  | ttgtctacaa  | atgatctaata | tcagccatta  | ttataaaaag  | aaggcttgag  | 81780 |
| tggaaaagag  | gctattgaaa  | gcatgcccac  | tattaataga  | tatggacaag  | attcaatttt  | 81840 |
| ttcagaaaata | gaagagctac  | tagagcatga  | tataaatatc  | attgccctat  | ttccagttat  | 81900 |
| taatgattcc  | aagaaaaaca  | gcactggaga  | tgaggctatc  | aatgcatcta  | atttgatgtg  | 81960 |
| tgaaactatc  | agtaacataa  | agaaacgatt  | ccctgaaata  | atcttaatat  | ctgatgttgc  | 82020 |
| tctagaccca  | tataccgata  | atgggcatga  | cggtatatta  | aaaaatgatt  | atgttgataa  | 82080 |
| tgatgagact  | ttagctgttt  | taaggaagca  | atcactaaca  | ttagctcagg  | ctgggacaga  | 82140 |
| cataattgca  | ccatcagata  | tgatggatgg  | aagaataggc  | tctataagag  | aggcattgga  | 82200 |
| cgaatggat   | tataaaaaata | caatccctact | atcatatgca  | gcaaaaatata | attcaaaagt  | 82260 |
| ttatgggtcca | tttagagatg  | ctgttaattc  | agcttcaagt  | ctaggcaagt  | cttcaaaaatc | 82320 |
| cagttatcaa  | atgtcaccaa  | aaaatatataa | tgaagctcta  | catgaagtgt  | ctatggatat  | 82380 |
| taatgaaggt  | gcagacatag  | tcatgggttaa | gccaggatgt  | ccttatctag  | atataaatttc | 82440 |

09847513 . 000001



|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| aaaagtaaaa  | gaaaccttca  | aagtaacctac | ctttgcatac  | caggtagtg   | gtgaatatag  | 82500 |
| catgcttaaa  | ctggcgattg  | ataaaggatg  | gcttgaaagc  | gatgttatgt  | tagaatcatt  | 82560 |
| aataagtttt  | aaaagagcgg  | gagctgatgc  | aattctaacg  | tacgcagcta  | aagaaaatttc | 82620 |
| caaggagata  | actaacaat   | gagcaatgtg  | atagaaattc  | gtgatgaaga  | aagctttaat  | 82680 |
| agcgacgtct  | taaattcaga  | aaaacctgta  | ttggttgatt  | tttgggctga  | gtggtgtgga  | 82740 |
| ccttgtaaac  | agcttgcacc  | aactgttgaa  | acagttgcag  | cagaaaaatc  | agaaacatta  | 82800 |
| aaggtttgca  | aatggatgt   | tgattcaaat  | agagagattg  | ctgcctaaata | tggaataaagg | 82860 |
| tcaatccctt  | cattaattat  | atttaaaaaac | ggagagcctg  | caggagtga   | agtaggtgct  | 82920 |
| ctaaccaaaac | aacaattaga  | ggactttata  | agtacagtag  | tttaactttg  | caaagacttc  | 82980 |
| tttgcatatt  | gcaagaaata  | ggattatcat  | ttcgacttct  | aaggctgaaa  | gccactcaaa  | 83040 |
| acaaaaaacc  | accttttcaa  | tcataaaaaac | aacaactaga  | acggaaaataa | ctaaatgaac  | 83100 |
| cttactgaaa  | ttaaaaataaa | accaataaat  | gaacttgtag  | atatagctac  | tgagcttggg  | 83160 |
| cttgaggatg  | ttggaaggct  | gaaaaagcaa  | gagataatat  | ttagaatatt  | taagcataag  | 83220 |
| gcttctgaag  | gtgttgatat  | ctatggtgcc  | ggagttcttg  | agattttaaa  | tgatgggtttt | 83280 |
| gggtttttgc  | gatccccaga  | aggctcttat  | tgcgctggcg  | aagatgatat  | ctatgtttca  | 83340 |
| ccaagccaaa  | taagaaaagt  | tagcctcagg  | aaggagact   | cagttgctgg  | gaagataaagg | 83400 |
| acccttaaa   | ataaagagcg  | ctatttttgc  | ttaatccaag  | ttgatactat  | taatgggtgaa | 83460 |
| gagccaagaa  | agactaaaaa  | caagattctt  | tttgaaaaatt | taactcctct  | ttttcccaat  | 83520 |

|             |            |             |             |             |             |       |
|-------------|------------|-------------|-------------|-------------|-------------|-------|
| gaaagactaa  | tccttgaaca | aggaacacagg | tctaatagaag | atcttttcac  | tcgaataaatt | 83580 |
| gatttgattg  | ctccaatagg | aaaagggtcag | cgtggactta  | tcgttttctcc | acccaaggct  | 83640 |
| ggtaaaaacct | taatgcttca | aagcatagct  | cattctatta  | aaagcaataa  | tccagaagta  | 83700 |
| gagcttatag  | ttcttttgat | tgatgaaaga  | cctgaagagg  | taacagagat  | gtcaagaact  | 83760 |
| gtaaaaggag  | aggttgtagc | tagtactttt  | gacgagccac  | ctactcgaca  | cgttcaagta  | 83820 |
| gcaaatatgg  | ttattgaaa  | agcaaaaaga  | cttgtagaac  | ataagaaaaga | tgtagttatc  | 83880 |
| ttattagatt  | ctattactcg | tcttggaaga  | gcataataact | cagttcagcc  | tgcatcagga  | 83940 |
| aagatatatga | gtggtggagt | tgactccaat  | gctcttgaaa  | ggccaaaaaag | gttttttgggt | 84000 |
| gctgctagaa  | atcttgaaga | gggtggaagt  | ctcactattc  | ttgctactgc  | tctagttgaa  | 84060 |
| acagggtcaa  | agatggatga | agttattttat | gaggaattca  | aagggtacggg | taatatggag  | 84120 |
| attcaccttg  | aaagaaaaat | agccgaaaaa  | agaatatatc  | ctgctattaa  | tattagaaga  | 84180 |
| tcgggaacaa  | gaagagaaga | tttacttact  | gctgaagatg  | aattacaaaag | aatgtgggtc  | 84240 |
| ttaaggaaaa  | ttttagacga | tatggaagat  | gctcagtcaa  | ttcagttcct  | aatagataga  | 84300 |
| ttgaaaatctc | ataaaaacaa | cgatgagttc  | tttacttcaa  | tgaaaaggggg | taatggcaag  | 84360 |
| aagactagat  | aaagtttttt | gccatatcaa  | tcatcatctt  | atcagttggg  | ctttgcgata  | 84420 |
| cgtgaatttc  | taattcttta | aactcatttt  | gacatctatc  | ctttatatatt | tctgaagcaa  | 84480 |
| caataaataat | tttttttctt | aaaactgcat  | cgtccagatt  | tgttattaaag | aatttaagag  | 84540 |

|             |             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tactaaaaatt | atagattaa   | aaaatttcat  | tattatctgt  | aacttttgggt | atttgctcca  | 84600       |       |
| gcaaatagat  | tacttcatag  | caaactatct  | catctagact  | agccttaagt  | ttttgttgaa  | 84660       |       |
| gaaaaccatt  | tgaatttttcg | ccgcaaaaata | aaagactctt  | tcctagaaaa  | tttttctctta | 84720       |       |
| ttaacttaag  | aattcccttca | gatgaatggc  | tttgtggaaa  | atgagatttt  | atgccacttt  | 84780       |       |
| caagaagt    | ttt         | gcggggcccaa | cagacaaaaa  | attgtgtgga  | agatcatctta | 84840       |       |
| aatcgaaaaa  | agatttttaa  | atatcaagtc  | cgtacgaagc  | agcagcttgg  | ctagtaaaaa  | 84900       |       |
| ttaagtttga  | atatgaatga  | atattttttaa | ttttatcgat  | agcgattttt  | gaagggtcga  | 84960       |       |
| ttgaattaat  | tttagaaaaga | taaatat     | tttt        | taagagctat  | ttcttgagct  | tcacataaaag | 85020 |
| atattagggt  | tcctgataaa  | tttttgggcc  | tcgtattaat  | tatcattttt  | taaaatttaa  | 85080       |       |
| gatttcgctc  | cttccaaaaat | aaactcatca  | gcaaataaaca | taatatctat  | tgcaatatct  | 85140       |       |
| tgaagggaag  | atatcttttc  | tttataaatt  | ttttcgttgc  | cttcataaga  | taatat      | ttttc       | 85200 |
| cctgatattt  | ttatttggcc  | atttttatct  | tcacataaaa  | tagctattgg  | agacaggcaa  | 85260       |       |
| gtaccttcca  | tggcagcaac  | aaaagatcgc  | tctgcactgg  | caagaattaa  | ttcttttaggg | 85320       |       |
| tcaccaatat  | tttcta      | aaag        | ctcaataata  | ttctttttat  | ctgacagaca  | ttcaattgct  | 85380 |
| atatatccct  | gagatgcaga  | aggtaacatt  | tttcaattg   | agaattcata  | tgaattttgt  | 85440       |       |
| tttaaacc    | ca          | gcctttttat  | agcagcctta  | gcaactacca  | gcccatcaaa  | cagaccatca  | 85500 |
| ttgagtttct  | ttattctagt  | agctatgttg  | cctcttactg  | gaatagtttt  | tatatcagggt | 85560       |       |
| cttaagttat  | taatttgtgc  | ttttcttctt  | ggaccagaag  | taccaatagt  | tgaatttcta  | 85620       |       |

09847513.000001

gaaattcag aaatgatatt cccatcccta gatagaagca ggtctttctgg agattcccta 85680  
ctaaaaacac ttattatttt aaacttttggg tcaagctttg ctggaacatc ttttaacta 85740  
tgcactgcaa tatctgcttc atcagcctct agagaggact caagtgtaga aatgaatagg 85800  
ccctttccac caatttcacg aagaggttta tctgttttgt ctcccttctga tgtcatagga 85860  
acaagctcaa ccttaatatt attgatctta gctaataact catctgcaac aaatttttgcc 85920  
tgatacattg ctagtctga ttgccttgta gctattctta ttttcatttt gttccaata 85980  
aaagagcctt cacatcgcca acactctttt ctttcaagat agtgaattcc tcaggaattt 86040  
ccaaggtagt aaatttacta tactcaaggt agattttaca agatgggttt atttcattct 86100  
ttcgaattat agatttttaga acttttaatt catattcttc gccaaaagga ggatcaagaa 86160  
gaattaaatc aaacgagat aagtcatttt ttttaatcca actaaatgca tccttaaaaa 86220  
aaactttaga tttatccttg atgccaagaa gttgaatatt ttttgctaac acagaatagt 86280  
tttttttatt aagttcaacg aacacaactt tttttgactg tctggatatt gcttcaatac 86340  
caagcgacc cgttcgggca aacaaatcaa gacatatata attttcaatc tcaaatgaa 86400  
gccaatataa aattatttct tttaatattat ttgaagttgg tcttaaggaa tccttaaaatt 86460  
cgaaaggat ttttttacct tttaaataac ctccagtgat tctgatattg tttttcatct 86520  
tttttgcaaa atgggtatgc ctatttatta aataaatata attamaata catttaaaact 86580  
aagttataat tcatccatgm gtccatacaag ataatttttag aaaagcaatg aggagctata 86640

|             |            |             |             |             |             |       |
|-------------|------------|-------------|-------------|-------------|-------------|-------|
| tttactctgt  | cagtgatg   | tcaaatgttg  | acgagaataa  | aaagtttagt  | gctataacag  | 86700 |
| tttcttcagt  | tacctctgtt | tctttagatc  | ctcctagttt  | gctcgtttgt  | atcaataaat  | 86760 |
| cagctggaat  | tcacaactca | ataaaaagaag | ggtcctctttt | ttgcataaac  | cttttaataa  | 86820 |
| aaaatcaaga  | agatatctt  | aatctatgca  | gttcatttaa  | gaccgaaggt  | gatagattta  | 86880 |
| atagtggtga  | ctgggattta | agcggcactc  | cgttttttaaa | aagtgccccag | gctaatatatt | 86940 |
| tttgtactgt  | tgatcaatta | atctcatacc  | acacacatac  | tattgtgatt  | ggtcatgtca  | 87000 |
| caaactctct  | tagcgatgag | aaaattaata  | cactgacata  | tgttgatggt  | agctatgaat  | 87060 |
| aaattttcaa  | aaaatgtatt | ttttatttta  | ataatactca  | actctttttt  | tcttgccctct | 87120 |
| aacatttttg  | ctagccagga | agagtggtgaa | gaaaagccaa  | gtgttttttat | tatctctcct  | 87180 |
| caagacgggt  | ttatctcaga | atctaataat  | gtaaaagtct  | tatttggatc  | aaaaaatatt  | 87240 |
| gaaataaaatc | cagctggcaa | aggtgagatt  | gcaaaaaata  | aatgttttgc  | aagcgggcat  | 87300 |
| caccatcttt  | tagttaatat | cgaagcattg  | ccagagagct  | ttattccttt  | tgacaagggt  | 87360 |
| tatttacatt  | ttggaggagg | tcaaaactgaa | acaattcttg  | atcttgatcc  | tggaacctat  | 87420 |
| tctctccaac  | taattcttgg | atcttatgtg  | cataattcaa  | aaatgcaggt  | aaataacttt  | 87480 |
| aaaggtcaag  | gacccttttt | atcagaaaaa  | ataacaatta  | cagttaatta  | gagattagac  | 87540 |
| cagttataac  | tttatctaga | tgttcatcta  | agttattagc  | aattatatatt | gcactagaaa  | 87600 |
| taggaccaga  | attatcgtca | taatcatcac  | caactacatt  | cacctttctt  | attatcccaa  | 87660 |
| caagaccatg  | caacatagac | caaagagtta  | tacatttaaa  | ggcaatcact  | tcctcaggct  | 87720 |

cgtctgctag attagcaaaa ctttttctca tattatcgta tgttccattt gcagatttaa 87780  
gaagatctgg gtaatcggca aagttaccaa cagctgttcc aaacatcaaa tcatatgtat 87840  
gtgcattttt taaaccaaac cctatatatt tgcttgcatt agtaacaagt tttttctttg 87900  
taattttttt tggattctca aaaaaaaca gctcatthaag ttttttaaa ccaacgggtg 87960  
caacagcggc atatacacat tctttttgtt caaaatgcct ataaggagca gtttgagaaa 88020  
caccactttc ctttgcaagg gatctaatac ttagtttagt gtaaccatct ctatcacaaa 88080  
gcctgcctgc gcattctata agttcttctc ttaagtttcc gtgatgatag ttattcataa 88140  
tttaatttta atgttgacac tgctaacata tatcatgtta ttatgtatac accgcataca 88200  
ttaagttgat tacatatata agtcaaatac agatatatga acataacaaa aataaataac 88260  
accgtcatag ctctttttatt aggcaatgcg tttctatcaa atcttgaggc tttagaagtt 88320  
cttgaggta aaatgcttga tgaatatgct gtaactagag aatttcctgg aaagctcatt 88380  
ccatcagacc agtctaagct agcattttgaa atacctggaa agataaaactc tattaatggt 88440  
gatacggag atgagggtcat cttaggggat gaacttgcct cattggatga tagagaagct 88500  
ttagcacaat taaatcaatc aaaggctaaa tttgatttag ctgaacaagt actagcaaga 88560  
tatataaatc tcagagcaga tgggcatatt tctattcaag atcttgataa ggctgaatct 88620  
gatcaaatag tagctaagtc gcagtatgat ttttatagag ttaaatttga gcaaaactaag 88680  
ttactagctc cttttaatgg agttattcaa aatagatttc tagatacagg atcagtaatc 88740

aatgcaggtg tccaagtttt agaaatttta ggctctagca atgttgaagc aagaatttct 88800  
attccaatga actatatgaa caaggttaag attggagatg agtatgaatt tgatatcaga 88860  
ggaatatcta caaaagctac gttagagaga ctggctccca tgtcaaccgg aggcctccaat 88920  
aataggtag caatttttag attgatacc tttttaatc caggatcaat agtaaagctc 88980  
aaactaagca tactgagaa agcaaaagggt acatgggttc caattaagtc actgtcccag 89040  
tccgaacaag gtatttgggc catatatacc attaacgagc aacaagtagt tgtagagat 89100  
cttggtgatg ttatttattt tgaagacgaa tatgcttttg tcagcggaac acttaataat 89160  
ggtgatttag taatttttag cggagctcaa aaaattattg aaggaaaaat aataaaaaata 89220  
aatgaatgtt attaattttc taatagaaa gcctaggata ttatttctaa cttagcatt 89280  
tatattactt tctggaattt cttctgggct ttcagttcca attcaagaaa accctgaact 89340  
ggctgagaga tggggaggtg ttcgtatttt tcttccccgg gcatacccag aaagaattga 89400  
aacagagata gtaaatgac tagaaatcaa acttagagaa gttgaagaaa tcgatgagct 89460  
tgaatcaatt attactcaag gtttttcaac aattgtagtt gaattaaatc aaagtgtacc 89520  
tcctatactt attgaagaga cttggtccaa ggttcaagac aagctcaatc aaatagttat 89580  
tcctcaagggt gcagaaaatat ttcttgatag aagcagtggt ccgcctatca ctgttcagta 89640  
tgctgtaacc tggaacggca gtggagatgc tccactaata atgatgtcca gactagcaag 89700  
ccagctaaaa agaaaaatata gctcaatagg ctcatctcat caaactgcaa ttttttggtga 89760  
aacagatgaa gaaatttttaa ttgaactaga ttcatcaaag ctatcttcgc ttggattatc 89820

|  |       |
|--|-------|
| atttcaagat atcgcaagtg ctattcaagc cctagatgca aaaaaaccta ttggtgtatc  | 89880 |
| ctcaaaacaac aattctgagc ttttatatag actcaaagat aatatacaga gcattcaaaa | 89940 |
| actctcagaa atacctatca aggttattaa taaatcagag atcatacagc tagatgatgt  | 90000 |
| ggcatttatt tcaaaaaatcc cggtttctcc tattgaagac atattcttgt ttaatggaaa | 90060 |
| tgtagttatt tctgttgctg gaaccggatc attttctcaa agagtccatg attatgtaga  | 90120 |
| acgcgcaaca attgtgttag atgagatgag agaaactctg ccgactgaga tcaatataga  | 90180 |
| tttagtttat gacgaatctg cttacacaac taaaaaatTT aatgagcctg taaaaagtTT  | 90240 |
| ttcattagca atattttttg ttttagcttt aagtcttttt tttcttgga ttagatcagc   | 90300 |
| aataattgtt actcttatcc tgccattttc tatttgcctg gttatgattg gttgtaggtt  | 90360 |
| tataggctta cccttgcata tgacatctat cactggaatt attattgcac taggattgct  | 90420 |
| tatagataat gggattattg ttgttgaaga ttataagaat agaagagcat ctggattaaa  | 90480 |
| tatcaatgat tcaatttcac aaggactaaa aaacttatgg gctccattag ctgctgcaac  | 90540 |
| agcaacaacc gtcttctctt ttcttcttat tgttactgga gaagggtcga gtattgaatt  | 90600 |
| cgtaggcggc atggcaatga cagtcattat gtctataaca tcttcattag ttttggcgtt  | 90660 |
| attaatgggt ccagttctga tgagttatat ggaaaaaaatt ccgttcttta aggatgtgga | 90720 |
| tattagcaag gaagggtata gaaatgaaaa aatccttaat aaatataggg ccttttttaa  | 90780 |
| ctgggcgttc ttagttccta gaagagcaat catgatatcg cttgcattgc ctgttcttagg | 90840 |



attctttctt tttaattctt tacctaaaga tttctttcct gctcaagata gagatatgtt 90900  
 tagagttaat atagaactgc cttctaacgc ctcatcactt actacaatgc agagagttaa 90960  
 ggaaattaga gaagatatct tagatagtga tttaatttca atagaaaaag attattcgtt 91020  
 tatcggcaga atgatgccta gagttttgat gaatgttgtt ggtggagaag aaaaacaagg 91080  
 atccaataat attgcgcagt ctgtattttt tgctactgat tattatgaaa tgattgaaaa 91140  
 ccttccagat ttatcaagaa gactgggttaa aaataaccct gacattatag ttaatataga 91200  
 tagtttctcg tctggcccc cggttttttc agatgttagt tatgtaattt ttggagatga 91260  
 tccagattta ctaaaatcac ttggtgagga gctagagcta attattaaca attctcctga 91320  
 tgtgagtctt acgaaatctg caacttcaa ctcaataacc aatgttgagt ttgaacttaa 91380  
 cagctcaaat atttcaactgt ctggtcaaaa tgccaattat ctgtaaatg aaatgtttac 91440  
 tgcaaacaaat ggaatatattg ttggcactat gttggattca acaaaagaaa taccagtcag 91500  
 gctgaaaagg ctgtctaata aaaacaatat tacgggaaat actagtttta taacaatgcc 91560  
 ctctcaagggt ggttttgagt attttgatag tttcggaaaa agctcactaa caaacaatac 91620  
 gtcaacaatt actaggcttg atggccaaag acaaaatgat gttgaggggt ggatttggac 91680  
 aggtacgctc ccgtctgcta ctgaaaaaatc tattaaaaaa gatgttaaag attttgaatt 91740  
 aagattgcca ataggctatt cattaaaaa acttggcgag gctgaaagca ggggcccaatc 91800  
 tcaagcctca ttatactctt cagcttttat gtatttcatt cttataatag taggcttggt 91860  
 tatggcgctt aattctttca gagaggctgg tctaattttg tctgttgcatt tcttatcaat 91920

|             |             |             |             |             |             |       |
|-------------|-------------|-------------|-------------|-------------|-------------|-------|
| tggactatcc  | tttcttgggt  | tatttatagg  | ccagcaaaat  | tatggattta  | ttggaactat  | 91980 |
| aagtgcaata  | gggttaattg  | gcttatcaat  | aaatgattca  | attattgtct  | tatctcacat  | 92040 |
| aaaagaagag  | gctgagaaga  | aatcactaac  | caaagctgag  | cttgttgaag  | ttgtttatcag | 92100 |
| gtctacacgt  | catataatca  | ctacctcttt  | gacaacacct  | ggtggttttc  | ttccacttat  | 92160 |
| ttttgcaagt  | gtattctttca | aaccgcttgc  | ttgggcaatg  | agcattggag  | tattaggcgc  | 92220 |
| gactattaca  | gccttattat  | atattcctgc  | aatgttttatg | ataatgagaa  | aggttaagta  | 92280 |
| ctagaacaac  | tttccgagca  | ataattttctt | cctagcttttt | ttacaataaag | tgatttcatga | 92340 |
| gtataggtct  | cgcaagaatc  | acattttaatc | atagagttat  | caatctttttt | tttatgaggt  | 92400 |
| cctgatatga  | gagatctgaa  | aaagaatacc  | gcagttatta  | tcaatataaa  | tatgaccaat  | 92460 |
| ggaatcaaaa  | gtaatatcga  | ttttaataaag | aacattttttt | agtcactaga  | tgctggctta  | 92520 |
| ctattttctta | taaaattcca  | agaccagcca  | gacttatctt  | cagattttttt | tgattgctca  | 92580 |
| tcttggtagt  | taatagaaat  | tatttttttta | gtatcttcaa  | gcaatttcaat | atatcccaaa  | 92640 |
| gattcatatg  | aagcctctaa  | gatcttaaga  | gctctataat  | tttcactaga  | gttaggaata  | 92700 |
| ttttcaatta  | cataaattgc  | tcttcttatt  | gctgctatat  | gcgcatcaac  | actaacgtaa  | 92760 |
| taatcagctg  | ccgcaagctc  | atttctttgca | atcataatttc | ttaaatagat  | atttctttgt  | 92820 |
| ttagcatatg  | tagaatattg  | gctatcagga  | aatcttggtta | agaattcagt  | tagtttctgaa | 92880 |
| aatgattcct  | tagctcctga  | gatatctcta  | tttgatagat  | cagtatctgt  | cattctttaca | 92940 |



|             |             |            |            |             |             |       |
|-------------|-------------|------------|------------|-------------|-------------|-------|
| taaagaaaat  | attggaaaact | actctgtgtt | ggatataaga | atagagacag  | gaagaaactca | 94080 |
| tcagattagg  | gtccatctat  | catcaaaaaa | actaccaata | attggagata  | aaacctatga  | 94140 |
| cccaagcagg  | tctattgcaa  | gagatacccc | tgaagagcta | attaatatata | tccgaggttt  | 94200 |
| tccaaggcag  | gcattacatg  | caacacacct | ctcattcaat | gaccaaaaaa  | caaataatat  | 94260 |
| tttttctttt  | gataattcca  | ctccaaatga | tatggaggaa | ctacttctag  | aattaagaaa  | 94320 |
| attgatctaa  | tagtaactaa  | aaacttgttt | tttgattaat | aaaataatat  | aaaccttatt  | 94380 |
| cctaagagtt  | ttttggtaag  | aaattgaaat | tatctggcgc | agacatgcta  | atgcaagcac  | 94440 |
| ttcatgatga  | agggtgtgag  | ctaattcttg | gctacccagg | tggagccgcg  | cttcatatct  | 94500 |
| atgatgcaat  | ttttagacaa  | gataaaatag | atcatatttt | agtaaggcat  | gagcaagggtg | 94560 |
| caaccatgc   | agcagatgga  | tattcaaggg | cgacaggtaa | gccaggagtt  | gtcttagtca  | 94620 |
| cttctggacc  | tggtgcaaca  | aatgctatta | caggaatcgc | gactgcattt  | atggattcca  | 94680 |
| taccaatggt  | agttatttca  | gggcagggtg | ctagccattt | aatagg tact | gatgcttttc  | 94740 |
| aagaaaactga | tatgattggt  | gtttcaagac | caattgttaa | gcatagctat  | acagttttta  | 94800 |
| atgctgaaga  | aataacctaa  | ataattaaag | aagcttttta | tgtcgcaact  | tcaggcagac  | 94860 |
| ctggacctgt  | tggtatagat  | atcccaaaag | acatgacagc | tccggataat  | cttttttgatt | 94920 |
| actcgtatcc  | tgaagaagcc  | aagataagat | catacaatcc | tccgattgag  | ccagaaaaaa  | 94980 |
| atcaaaataga | tagagcagtc  | gaagctatat | tgatatcaaa | aaaaccagtt  | atatatgctg  | 95040 |

09847513.000001

|             |             |             |            |             |             |       |
|-------------|-------------|-------------|------------|-------------|-------------|-------|
| gtggtggggc  | aattgctagt  | aatgccgaaa  | agaattact  | tgaacttaat  | gaaattattg  | 95100 |
| atgctcctgt  | tacaaatact  | ttaatgggat  | tgggtattta | cctgctagt   | catcatagat  | 95160 |
| ttcttgggat  | gttagggatg  | catggaacat  | atcaggcaaa | tatggcaatg  | cataatgcag  | 95220 |
| acttaataat  | tgctattggc  | gccagatttg  | atgacaggat | taccaataaa  | ccatcaaaat  | 95280 |
| ttgcacctaa  | tgccaaaagt  | gttcatctag  | atgttgatca | ctcatctgta  | tcaaaaaatta | 95340 |
| tagaagcaaa  | tgtagctgtt  | tttgggcaag  | taaaaaatc  | cttaaaaatta | ataaaagaaa  | 95400 |
| ctcttgaaaa  | aaaattagac  | tcttacgatt  | ctttcgctct | tcagccttgg  | cacgatcaga  | 95460 |
| taaaagaatg  | gaaatcacta  | catggtttta  | attatgagct | ttataaagat  | gaatctgatg  | 95520 |
| atcatcccat  | tttaccgccag | gctgtagtcc  | agcatgtcca | tgagattaca  | aatggggaag  | 95580 |
| cataatgtgac | ttccgatgtt  | ggtcagcadc  | aaatgtttgc | tgctcaatat  | tatcatttttg | 95640 |
| ataagcctag  | aagatggatc  | aattctgggtg | gtctaggaac | tatgggtttt  | ggtttgccag  | 95700 |
| cagcaaatggg | tgtaaaaactc | gctttttccaa | aagatgaggt | tgtttgcatt  | actggtgagg  | 95760 |
| gtagtatcca  | aatgtgcadc  | caagaattgt  | ctacatgtct | tcaatataat  | ctcccaataa  | 95820 |
| aaataattaa  | tatcaataac  | gaagctcttg  | gtatggttta | acaatggcag  | gatatgaatt  | 95880 |
| atggaggaag  | gcactctgaa  | agtacctatc  | aaaactcgtt | accggatttt  | ataaaaactgg | 95940 |
| ctgaatcata  | tggtcatata  | ggaattaaaa  | ttacaaaaaa | ttctgattta  | agtgaaggct  | 96000 |
| taaaaaaagc  | ttttgaaatg  | aaagataaac  | ttgtctttgt | tgataatttat | gtagatcctt  | 96060 |
| cagagcatgt  | ttatccaatg  | caggttgcaa  | atggcagcct | agaaaaatatg | tggctatcaa  | 96120 |

aggatgaaca aacatgatta aaagaaaact aattttaatt atggaaaata aaccaggagc 96180  
 tctagtaaga gtagttggac tgtttcatca aagaggctac aacattgaaa cccttcatgt 96240  
 agatactgtt aaagactttt ctacttacaa atcgatatgt aaaaaaaact tgaaacccaaa 96300  
 tttgaggata atcaaatatc tagactgacc atagaaacaa tggtttcaga tgaccttatg 96360  
 aggcagattt tgagacagct caataaatta atagatgtta tagctgtaag caatgaagag 96420  
 acaacctatt taaaaggagt attattagat gaaaatttat tatgacgaag atgcaaacat 96480  
 agaaattatt aaagggatga atgtctcaat aattggctat ggatctcaag ggaatgctca 96540  
 tgcaaaatat cttcatgaat ctggtgtaag tgttactgta ggtttaagag aagggtcttc 96600  
 ttcatgggca aaggcagaag aagcaggctt aaaagttcaa acagttgctg attcggtaat 96660  
 ccaagcagat ttggttatga ttttggcacc agatgaattc caaaaaata tatatgaac 96720  
 cgaaatcaag ccaaacttaa aaaccagtgc aattcttgca ttgacacatg gctttaatat 96780  
 tcatttttgaa aaaatagttc ctgaagcaac taacagcgtta attatgattg caccaaaaag 96840  
 tccaggccat actgttagaa gtacttatac caatggtgga ggcgttccat ctctcatagc 96900  
 tataatatgaa gatgcctttaa gtgatgaaga ttattcagca aaagatgtag ctctatctta 96960  
 tgcaaaaagca aatggcggca cgagggctgg tgttcttgaa acatctttaa aagaagaaac 97020  
 agaaacagat ttatttggcg aacaagcagt tttatgtggc gggcttaccg ctttaattaa 97080  
 agctgggttt gaaactctag ttgaggcagg ctacagtga gagatggcat attttgaatg 97140

|             |             |             |             |             |              |       |
|-------------|-------------|-------------|-------------|-------------|--------------|-------|
| ccttcatgaa  | acaaaaactaa | tcacagactt  | aattcaagaa  | ggtggcattg  | ctaatatgca   | 97200 |
| ttactcaata  | tcaaataactg | ctgagtatgg  | tgattatgtg  | agtggaacca  | aagtaattac   | 97260 |
| tagcgatacc  | aaaaaagcaa  | tgaaaggaat  | actagaaaaat | atacagtcag  | gaaaattcgc   | 97320 |
| agatgacttc  | ttgaatgact  | gtcgacaaaag | caatgatggc  | actggcggac  | ctgttatgaa   | 97380 |
| aagcaataga  | gaagctacaa  | aaattcatcc  | aatagagtcg  | gtaggggctg  | agctaaggtc   | 97440 |
| taaaatgaag  | ttcctaaatt  | cacaaaaaatt | ggtggataaa  | gaaattaatt  | aaaattaatt   | 97500 |
| aaaaaaaaag  | tatcttcttc  | ggaataaactt | cgttaagata  | cgcgtccgca  | caaagaggcg   | 97560 |
| gttgttcttt  | aaaaatatatt | ggttactcgt  | gtgggtgttc  | aaaatacagag | aaaaaataat   | 97620 |
| ttagattttt  | tataaaaaatc | aacaaaaacat | gatattaatt  | gaagagtttg  | atcatggctc   | 97680 |
| agattgaacg  | ctggcggtag  | gcttaacaca  | tgcaagtcgt  | gcgagaaagt  | atcttcggat   | 97740 |
| atgagtagag  | cggcggacgg  | gtgagtaacg  | cgtaggaatc  | tacctagtag  | aaggggatag   | 97800 |
| ccccgggaaa  | cccggaattaa | taccgtatac  | ctccttcggg  | agaaagaagg  | cctctctttg   | 97860 |
| aagctttcgc  | tactagatga  | gcctgcgtaa  | gattagcttg  | ttggtgaggt  | aaaggctcac   | 97920 |
| caaaggcgacg | atctttagct  | ggtctgagag  | gacgatacgc  | cacattggga  | ctgagacacg   | 97980 |
| gcccagactc  | ctacgggagg  | cagcagtggg  | gaatattgga  | caatggggcg  | aagcctgac    | 98040 |
| cagccatacc  | gcgtgtgtga  | agaaggcctt  | cgggttgtaa  | agcactttaa  | gcaggggagaa  | 98100 |
| aaagttataa  | gttaataacct | tataaccctg  | atgttacctg  | cagaataagc  | accgggctaatt | 98160 |
| tccgtgccag  | cagccgcggt  | aatacggaag  | gtgcaagcgt  | taatcggaat  | tactgggcgt   | 98220 |

|            |              |             |             |             |             |       |
|------------|--------------|-------------|-------------|-------------|-------------|-------|
| aaagcgcgcg | tagtggtgtt   | gttaagtgg   | atgtgaaagc  | cctggggtca  | acctaggaac  | 98280 |
| tgcatccaaa | actaactcac   | tagagtacga  | tagagggagg  | tagaattcat  | agtgtagcgg  | 98340 |
| tggaatgcgt | agatatattatg | aagaataacca | gtggcggaagg | cggcctcctg  | gactctgtact | 98400 |
| gacactgagg | tgcgaaaagcg  | tgggtagcga  | acaggattag  | ataccctggt  | agtccacgcc  | 98460 |
| gtaaacgatg | acaactagct   | gttgggagac  | tatgtctttc  | agtggcgag   | ctaacgccttt | 98520 |
| aagttgtccg | cctgggggagt  | acggccgcaa  | ggctaaaact  | caaatgaatt  | gacgggggacc | 98580 |
| cgcacaagcg | gtggagcatg   | tggtttaatt  | cgatgcaacg  | cgaaaaaacct | tacctactct  | 98640 |
| tgacatactt | ggaggctctt   | gtaatgagag  | tgtgcctttt  | ggaaccaaga  | tacagggtgct | 98700 |
| gcatggctgt | cgtcagctcg   | tgtcgtgaga  | tgttccgtta  | agtcggataa  | cgagcgcaac  | 98760 |
| ccttaccctt | atttgccagc   | gattcggtcg  | ggaactataa  | ggggactgcc  | ggtgataaac  | 98820 |
| cggaggaagg | tgaggacgac   | gtcaagtcat  | catggccctt  | acgagtaggg  | ctacacacgt  | 98880 |
| gctacaatgg | gagatacaga   | cggacgcctaa | gccgcgaggt  | ggtgctaatac | ctaaaaagtc  | 98940 |
| tttcgtagtc | cggattggag   | tctgcaactc  | gactccatga  | agtcggaatac | gctagtaatac | 99000 |
| gcggatcagc | atgccgcggt   | gaatacgttc  | tcgggtcttg  | tacacacccg  | ccgtcacacc  | 99060 |
| atggaagtgg | attgcaccag   | aagtagatag  | tctaacctta  | gggagggcgt  | ttaccacggt  | 99120 |
| gtgcttcatg | actgggggtga  | agtcgtaaca  | aggtagccgt  | aggggaacct  | gtggctggat  | 99180 |
| cacctcctta | acgataaaatc  | gcgtttttaaa | cggccacacg  | agtaatcaaa  | tattaaaaaa  | 99240 |



aagaacattt agatatgtaa aatcatttggt atgtaatttt ctagtgtata catttatgta 99300  
tacataagat cactgcaatt aaaaagtaac atatgcattt atgtgtatgt taaaaaagta 99360  
attaatatat ttattttaagt tacttctaaa aatgaagata aaacttcaaa aaaaatatgt 99420  
aacctttttt aaggttatat gatcaagtaa aggaagagca caaggcggat gccttggcag 99480  
cataaggcga tgaaggacgt aataacctgc gataagcctc ggggagctgg taaataagct 99540  
tcgatccgag gattccgaa tgggaaaacc caatacacat aagtgtatta tcttatactg 99600  
aatacatagg tataagaggc aaacctaggg aactgaaaca tctaagtacc tagaggaaaa 99660  
gaaatcaaca gagattccgg tagtagcggc gagcgaaaacc ggaccagccc ttaagcttat 99720  
tttagtccag caaaatatc tggaaagtth agccatagta ggtgatagcc ctgtatgcga 99780  
aagactaatt taagtgaat cgagtaggtc gggacacgag aaatcttgac tgaacatggg 99840  
gggaccatcc tccaaggcta aatactctat gctgaccgat agtgaaccag taccgtgagg 99900  
gaaaggcgaa aagaaccccg gcgagggggag tgaaatagaa cctgaaacct tgtgcttaca 99960  
agcagtcgga gcagacttgt tctgtgacgg cgtacctttt gtataatggg tcaacgactt 100020  
aatttcagta gcaagcttaa ccatttaggg taggcgtagg gaaaccgagt cttaataggg 100080  
cgctcagttg ctggaattag acccgaaaacc ggggtgatcta tccatggcca gtgtgaagg 100140  
cgagtaacat cgactggagg cgcgaaccca cttatgttga aaatgaggg gatgagctgt 100200  
ggataggagt gaaaggctaa tcaaacccgg agatagctgg ttctcttcga aaactattta 100260  
ggtagtgcct cgtgtattac tgtagggggg agagcactgt ttcggctagg gggatcatccc 100320

gacttaccaa accgatgcaa actccgaata cctacaagta tgagcacggg agacagactg 100380  
cgggtgctaa cgtccgtagt cgagagggaa acaaccacaga ctgtcagcta aggtcccaaa 100440  
ttatgattaa gtgggaaaca atgtgggaag gcacaaacag ctaggaggtt ggcttagaag 100500  
cagccatcct ttaaagaaa cgtaatagct cactagtcga gtcggccctgc gtggaagata 100560  
taacggggct aaatcataaa ccgaagctac agatcttataa ttattttaag atggtagaag 100620  
agcgttctgt aagcggttga aggtaagctg agaggcgaac tggacgtatc agaagtgcga 100680  
atgttgacat gagtaacgat caaagagggtg aaaaacctct tcgccgaaaa accaagggtt 100740  
cctgtccaac gctaatacgag gcagggtgag gcggccccta aggcgagggc gaaagccgta 100800  
gtcgatggga aacaggttaa tattcctgta ctttttataa ctgcgatggg gtgacggaga 100860  
aggtagact agcacggcg cggttgtccg tgttcaagggt tgtaggctgg tgttctaggt 100920  
aaatccggaa cgctaaggct gagaactgat aacgaccact ctacgagtgg gaagtagtcg 100980  
ataccatgct tccaggaaaa acctctaagc ttcaggttat aagaaaccgt accctaaacc 101040  
gacacagggt gtaggtcga gtagaccaag gtgtttgaga gaactatggt gaaggaaacta 101100  
ggcaaaaatag caccgtaact tcgggagaag gtgcgccgcg tttggtgatg agacttgctc 101160  
tctaagctga acgtggtcga agataaccagg tggctgcgac tgtttactaa aaacatagca 101220  
ctctgcaaac tcgtaagagg aagtatagggt tgtgacgcct gcccggtgcc ggaagggttaa 101280  
ttgatgggt tagcttatgc gaagctcttg atcgaagccc cggtaaacgg cggccgtaac 101340

tataacgggtc ctaaggtagc gaaatttcctt gtcgggtaag ttccgacctg cacgaatggc 101400  
gtaacgatgg ccacactgtc tccaccatag actcagtgaa attgaaatcg ctgttaagat 101460  
gcagtgtagc cgcagctaga cggaagacc ccgtgcacct ttactatagg ttgcgactgg 101520  
actttgacct tacttgtgta ggataggtag gagactttga agcagagacg ccagtctttg 101580  
tggagtcatc cttgaaatac cacccttgta agattgaagt tctaaccctag gtccattatc 101640  
tggatcaggg acagtgcgtg ctgggtagtt tgactggggc ggtctcctcc taaagagtaa 101700  
cggaggagta cgaaggtatc cttatcacgg tcggacatcg tgaggtaagt ataaaggcag 101760  
aaggatgctt gactgcgaga tcgacggatc gagcaggtag gaaactaggc cttagtgatc 101820  
cggtggttct gaatggaagg gccatcgctc aacggataaa aggtacgccg gggataaacg 101880  
gctgataccg cccaagagtt catatcgacg gcggtgtttg gcacctcgat gtcgggctcat 101940  
cacatcctgg ggctggagca ggtcccaagg gtatggctgt tcgccattta aagtggtagc 102000  
cgagctgggt ttagaacgtc gtgagacagt tcggtcccta tctgctgtgg gcgtttggag 102060  
atttgagggg agctgattct agtacgagag gaccgaattg gacgaacctc tgggtgttccg 102120  
gttgtcacgc cagtggcatt gccgggtagc tatgttcgga aaggataaacc gctgaaaagca 102180  
tataagcggg aagcctctcc caagattaaa tctcccagag actttatgtc tcctaaagag 102240  
tcgtcataga ctatgacgtt gataggcaag atgtgtaagc gctgcgaggc gttgagctaa 102300  
cttgtactaa taactcgtga ggcttgatca tgtaacctta agcaagggtc ataatttgag 102360  
taaaacatbg tagtgagaat taaaaataa aaagttacat accagtttgc ctgatgacaa 102420

tagcaacttg gaaccacctg atcccatctc gaactcagaa gtgaaacgag ttaacgccaa 102480  
tggtagtgc gggctctccct gtgtgagagt aggaatcgt caggcttttt tctttaaggc 102540  
ttccagttta ctggaagcct ttttttttat ctcaagtata ataccagat gattattgga 102600  
ttaacaggag gcattggttc tggcaaatct gccgctgcag acttctttat tgatttaggt 102660  
atatcagtct tagatgcaga tcaagttgct aaagaagctt tatctacaaa ttctcctgga 102720  
tatactgatt ttatttctca atttggtgaa gtgtatttaa ataataatcg tgaggttgat 102780  
aggctgaaat tgcgcgaact tatttttttca aatccttcaa aaaaaaggga tcttgagaat 102840  
attattcatc ccatagttcg gtctgctatt agtaatttta ttattacatc aacatcacca 102900  
tattctattg ttatggtgcc actcattttt gaaacaaatt catataaaa ttacgataag 102960  
attattactg ttgactgtga tttagaactt cagatagtaa gggcctcaag tagagatgct 103020  
caaaataaat cgcagattaa gaatattatt aataagcaag cctctagaga ggagaggcta 103080  
agtattttctg atgacgtact tatcaataac agcaccttat ctgatctaaa aaaacaagtt 103140  
aatgttttac atactaaata tatggagtta ttaaatgagt agttgcccta gatgtgaaaa 103200  
acctgtcaaa ctttctactg acaatattta tagacctttc tgctctgaga aatgcaaaact 103260  
tatagatttt ggtgattggg ctaatgaaga taataaaaatt tcaagaccaa ttcaatctga 103320  
agatttttac gaagattaaa tttaagaaaag tctccattca ccactttcaa tcatcggttt 103380  
ggctttttta tacttcattt ctgcgtatc ttgaccgtta gtaattttta caagatcatt 103440

tctgcctaac ttaggttcgt ttcttgtcac agtacttggt ttactatag gagtggcttg 103500  
tttttcattt tgaagatat cagaattaat ttctctttt tccaatttta attcttgaga 103560  
actattttgt ttgtttatgc tctctagctc agattctgtc gatatattgta atgagaaaaag 103620  
tattcttata gtttcaacat ctatttccga aagcatagat tcaaacattg aatatgcctc 103680  
tcttttgaat tcatttttttg gatttttttg agcataagcc cttagaccac cactatttct 103740  
taaatgatct atttctgata aatgctcttt ccaatggaca tcaagtactt gcagcataac 103800  
ctgcttctca agaagcaatc tattttcacc aagatcacta aacttttttg aatatttatt 103860  
ttttgctgc agaacaattt ctctggcaat tgtatttggt acaagttttt tattactatt 103920  
aattttattg gctatatctg ttctagtcc atagctctct tttaataat catcgagctg 103980  
cctacttttc cattgagact ctattgattc ttcaggtaac tataaattag atattccttt 104040  
gaattgctgc tcgatgagtg actcaatggt actgctaata tcttcctctt ctagtaattg 104100  
attccttaaa gaatatatag ctgtgtcttg atcatttgat acatcgtcgt actctaaaag 104160  
atttttcctt gcatcaaaagt ttctgtcttc aattcttttt tgcgcatttt ctattcctct 104220  
tgaaagcatt ttagcttcaa tatgatcatc tcccatgccc agcctttcaa aagtagccct 104280  
tcttccatcc gaaataaaaa gtctcaagag atcgtcttct aaagataaga agaatttttg 104340  
ataacctgga tcaccttgtc tgcctgatct ccccttaac tgattatcta ttcttcttga 104400  
ttcatgcctt tctgttccaa gtatatgaag ccaccagat tctataactt ttttattatt 104460  
ttctttccac tctatatcgc ttgtgtcttc ttttttccg ccaagaacta tatccgtccc 104520

ccttcctgcc atattcgtag caatagttac cattccaggt ttacctgcat tggcaattat 104580  
ctctgcttct ttttcgtgat gcttggcatt taaaatttga tgaggtattt ttttgttatt 104640  
taaatatgct gatacttctt ctgaagattc aaccgaaact gttccaacaa gaatgggaga 104700  
agattttttt cttaattgtt caatttcttc aattagagct ttatatattcg attctgttgt 104760  
aagaaaatact aagtcattaa gatcagctct agccatagga acatttgttg ggatgatgat 104820  
gacatttagg ccatagattt gactaaactc tactgcttca gtatctgctg ttcctgtcat 104880  
cccagaaaagt tttttaaata atctaaaaaa gtttttggaat gtggtggatg ctagtgtttg 104940  
agactctctt tggatagcaa cattttcttt gcattccagt gcctgggtgaa caccttcaact 105000  
cattcttctt ccgggcattg ttctacctgt atgctcatca atcaaaaagaa cctcacccgtt 105060  
cctaaccacaa taatccacat tcttttttaa taagaagctt gctctaagtg ttgcttgaac 105120  
aaatttcata atttttaaat tagaaacaga gtaagcccat ctgaggctcc aagccgattc 105180  
cagc 105184

<210> 2  
<211> 29  
<212> DNA  
<213> pcr primer  
<220>  
<221> misc\_feature  
<222> (1) .. (29)  
<223> Oligonucleotide primer for amplifying the proteorhodopsin gene. F

orward primer

<400> 2 29  
accatgggta aattattact gatattagg

<210> 3  
<211> 24  
<212> DNA  
<213> pcr primer  
  
<220>  
<221> misc\_feature  
<222> (1)..(24)  
<223> Oligonucleotide primer for amplifying the proteorhodopsion gene.  
Reverse primer

<400> 3 24  
agcattagaa gattctttaa cagc

<210> 4  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium  
  
<220>  
<221> CDS  
<222> (1)..(750)  
<223> Light-driven proton pump that is active when expressed in E. coli  
, with retinal. An additional three nucleotides are incorporated  
to native sequence (31A08) via pcr primer (DNA residues 4-6, ggt  
) , adding a new restriction site for cloning/expression

```

<400> 4
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu 5 Ile Leu Gly Ser Val 10 Ile Ala Leu Pro Thr 15
1
ttt gct gca ggt ggt ggt gac gct gct gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val 20 25 30
tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe 35 40 45
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50 55 60
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met 65 70 75 80
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr 85 90 95
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu 100 105 110
att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta 384
Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu 115 120 125

```



cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
  
 aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
  
 att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
  
 tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr  
 210 215 220  
  
 aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240  
  
 aat gtt gct gtt aaa gaa tct tct aat gct 750  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 5  
 <211> 250

09847513.0000001

<212> PRT  
 <213> Naturally occurring gamma proteobacterium  
  
 <400> 5  
  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15  
  
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30  
  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
130 135 140

Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Lys Ser Ala Cys  
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 6



|   |     |
|---|-----|
| ggt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act gta<br>Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr Val<br>50 55 60        | 192 |
| tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg aga<br>Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met Arg<br>65 70 75 80     | 240 |
| ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac att<br>Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr Ile<br>85 90 95        | 288 |
| gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta att<br>Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu Ile<br>100 105 110     | 336 |
| ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta cta<br>Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu Leu<br>115 120 125     | 384 |
| ggt ggt tct ctt gtt atg ctt gtt ttt ggt tac atg ggt gaa gca gga<br>Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala Gly<br>130 135 140     | 432 |
| atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg gta<br>Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp Val<br>145 150 155 160 | 480 |
| tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt aat<br>Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys Asn<br>165 170 175     | 528 |
| act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat att<br>Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr Ile                    | 576 |

|   |     |     |     |
|---|-----|-----|-----|
| 180   | 185 | 190 |     |
| atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt tac |     |     | 624 |
| Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly Tyr | 200 | 205 |     |
| 195   |     |     |     |
| ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat aac |     |     | 672 |
| Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr Asn         | 215 | 220 |     |
| 210   |     |     |     |
| ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg aat |     |     | 720 |
| Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp Asn | 230 | 235 |     |
| 225   |     |     |     |
| ggt gct gtt aaa gaa tct tct aat gct                             |     |     | 747 |
| Val Ala Val Lys Glu Ser Ser Asn Ala                             | 245 |     |     |
|   |     |     |     |
| <210> 7   |     |     |     |
| <211> 249   |     |     |     |
| <212> PRT   |     |     |     |
| <213> Naturally occurring gamma proteobacterium                 |     |     |     |
| <400> 7   |     |     |     |
| Met Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr Phe | 5   | 10  | 15  |
| 1   |     |     |     |
| Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val Ser | 20  | 25  | 30  |
|   |     |     |     |
| Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe Phe |     |     |     |

35

40

45

Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr Val  
50 55 60

Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met Arg  
65 70 75 80

Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr Ile  
85 90 95

Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu Ile  
100 105 110

Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu Leu  
115 120 125

Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala Gly  
130 135 140

Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp Val  
145 150 155 160

Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys Asn  
165 170 175

Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Tyr Ile  
180 185 190

Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly Tyr  
195 200 205

Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr Asn  
210 215 220

Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp Asn  
225 230 235 240

Val Ala Val Lys Glu Ser Ser Asn Ala  
245

<210> 8  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium  
  
<220>  
<221> CDS  
<222> (1)..(750)  
<223> proteorhodopsin variant from clone EBAC40

<400> 8  
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15 48





|   |     |     |     |     |
|---|-----|-----|-----|-----|
| 145   | 150 | 155 | 160 |     |
| gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt     |     |     |     | 528 |
| Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys     | 165 | 170 | 175 |     |
| aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat     |     |     |     | 576 |
| Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr     | 180 | 185 | 190 |     |
| ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt     |     |     |     | 624 |
| Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly     | 195 | 200 | 205 |     |
| tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat     |     |     |     | 672 |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr             | 210 | 215 | 220 |     |
| gac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg     |     |     |     | 720 |
| Asp Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp     | 225 | 230 | 235 | 240 |
| aat gtt gct gtt aaa gaa tct tct aat gct                             |     |     |     | 750 |
| Asn Val Ala Val Lys Glu Ser Ser Asn Ala                             | 245 | 250 |     |     |
| <br><210> 9   |     |     |     |     |
| <211> 250   |     |     |     |     |
| <212> PRT   |     |     |     |     |
| <213> Naturally occurring gamma proteobacterium                     |     |     |     |     |
| <br><400> 9   |     |     |     |     |
| <br>Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr |     |     |     |     |

|      |     |     |     |
|------|-----|-----|-----|
| 1    | 5   | 10  | 15  |
| Phe  | Ala | Ala | Gly |
| 20   | Gly | Gly | Asp |
| 25   | Leu | Ala | Ser |
| 30   | Asp | Tyr | Thr |
| 35   | Gly | Gly | Val |
| 40   | Ala | Ala | Leu |
| 45   | Ala | Ser | Thr |
| 50   | Val | Thr | Val |
| 55   | Ala | Leu | Phe |
| 60   | Ala | Ser | Thr |
| 65   | Val | Thr | Val |
| 70   | Ala | Ser | Thr |
| 75   | Ala | Ser | Thr |
| 80   | Ala | Ser | Thr |
| 85   | Ala | Ser | Thr |
| 90   | Ala | Ser | Thr |
| 95   | Ala | Ser | Thr |
| 100  | Ala | Ser | Thr |
| 105  | Ala | Ser | Thr |
| 110  | Ala | Ser | Thr |
| 115  | Ala | Ser | Thr |
| 120  | Ala | Ser | Thr |
| 125  | Ala | Ser | Thr |
| 130  | Ala | Ser | Thr |
| 135  | Ala | Ser | Thr |
| 140  | Ala | Ser | Thr |
| 145  | Ala | Ser | Thr |
| 150  | Ala | Ser | Thr |
| 155  | Ala | Ser | Thr |
| 160  | Ala | Ser | Thr |
| 165  | Ala | Ser | Thr |
| 170  | Ala | Ser | Thr |
| 175  | Ala | Ser | Thr |
| 180  | Ala | Ser | Thr |
| 185  | Ala | Ser | Thr |
| 190  | Ala | Ser | Thr |
| 195  | Ala | Ser | Thr |
| 200  | Ala | Ser | Thr |
| 205  | Ala | Ser | Thr |
| 210  | Ala | Ser | Thr |
| 215  | Ala | Ser | Thr |
| 220  | Ala | Ser | Thr |
| 225  | Ala | Ser | Thr |
| 230  | Ala | Ser | Thr |
| 235  | Ala | Ser | Thr |
| 240  | Ala | Ser | Thr |
| 245  | Ala | Ser | Thr |
| 250  | Ala | Ser | Thr |
| 255  | Ala | Ser | Thr |
| 260  | Ala | Ser | Thr |
| 265  | Ala | Ser | Thr |
| 270  | Ala | Ser | Thr |
| 275  | Ala | Ser | Thr |
| 280  | Ala | Ser | Thr |
| 285  | Ala | Ser | Thr |
| 290  | Ala | Ser | Thr |
| 295  | Ala | Ser | Thr |
| 300  | Ala | Ser | Thr |
| 305  | Ala | Ser | Thr |
| 310  | Ala | Ser | Thr |
| 315  | Ala | Ser | Thr |
| 320  | Ala | Ser | Thr |
| 325  | Ala | Ser | Thr |
| 330  | Ala | Ser | Thr |
| 335  | Ala | Ser | Thr |
| 340  | Ala | Ser | Thr |
| 345  | Ala | Ser | Thr |
| 350  | Ala | Ser | Thr |
| 355  | Ala | Ser | Thr |
| 360  | Ala | Ser | Thr |
| 365  | Ala | Ser | Thr |
| 370  | Ala | Ser | Thr |
| 375  | Ala | Ser | Thr |
| 380  | Ala | Ser | Thr |
| 385  | Ala | Ser | Thr |
| 390  | Ala | Ser | Thr |
| 395  | Ala | Ser | Thr |
| 400  | Ala | Ser | Thr |
| 405  | Ala | Ser | Thr |
| 410  | Ala | Ser | Thr |
| 415  | Ala | Ser | Thr |
| 420  | Ala | Ser | Thr |
| 425  | Ala | Ser | Thr |
| 430  | Ala | Ser | Thr |
| 435  | Ala | Ser | Thr |
| 440  | Ala | Ser | Thr |
| 445  | Ala | Ser | Thr |
| 450  | Ala | Ser | Thr |
| 455  | Ala | Ser | Thr |
| 460  | Ala | Ser | Thr |
| 465  | Ala | Ser | Thr |
| 470  | Ala | Ser | Thr |
| 475  | Ala | Ser | Thr |
| 480  | Ala | Ser | Thr |
| 485  | Ala | Ser | Thr |
| 490  | Ala | Ser | Thr |
| 495  | Ala | Ser | Thr |
| 500  | Ala | Ser | Thr |
| 505  | Ala | Ser | Thr |
| 510  | Ala | Ser | Thr |
| 515  | Ala | Ser | Thr |
| 520  | Ala | Ser | Thr |
| 525  | Ala | Ser | Thr |
| 530  | Ala | Ser | Thr |
| 535  | Ala | Ser | Thr |
| 540  | Ala | Ser | Thr |
| 545  | Ala | Ser | Thr |
| 550  | Ala | Ser | Thr |
| 555  | Ala | Ser | Thr |
| 560  | Ala | Ser | Thr |
| 565  | Ala | Ser | Thr |
| 570  | Ala | Ser | Thr |
| 575  | Ala | Ser | Thr |
| 580  | Ala | Ser | Thr |
| 585  | Ala | Ser | Thr |
| 590  | Ala | Ser | Thr |
| 595  | Ala | Ser | Thr |
| 600  | Ala | Ser | Thr |
| 605  | Ala | Ser | Thr |
| 610  | Ala | Ser | Thr |
| 615  | Ala | Ser | Thr |
| 620  | Ala | Ser | Thr |
| 625  | Ala | Ser | Thr |
| 630  | Ala | Ser | Thr |
| 635  | Ala | Ser | Thr |
| 640  | Ala | Ser | Thr |
| 645  | Ala | Ser | Thr |
| 650  | Ala | Ser | Thr |
| 655  | Ala | Ser | Thr |
| 660  | Ala | Ser | Thr |
| 665  | Ala | Ser | Thr |
| 670  | Ala | Ser | Thr |
| 675  | Ala | Ser | Thr |
| 680  | Ala | Ser | Thr |
| 685  | Ala | Ser | Thr |
| 690  | Ala | Ser | Thr |
| 695  | Ala | Ser | Thr |
| 700  | Ala | Ser | Thr |
| 705  | Ala | Ser | Thr |
| 710  | Ala | Ser | Thr |
| 715  | Ala | Ser | Thr |
| 720  | Ala | Ser | Thr |
| 725  | Ala | Ser | Thr |
| 730  | Ala | Ser | Thr |
| 735  | Ala | Ser | Thr |
| 740  | Ala | Ser | Thr |
| 745  | Ala | Ser | Thr |
| 750  | Ala | Ser | Thr |
| 755  | Ala | Ser | Thr |
| 760  | Ala | Ser | Thr |
| 765  | Ala | Ser | Thr |
| 770  | Ala | Ser | Thr |
| 775  | Ala | Ser | Thr |
| 780  | Ala | Ser | Thr |
| 785  | Ala | Ser | Thr |
| 790  | Ala | Ser | Thr |
| 795  | Ala | Ser | Thr |
| 800  | Ala | Ser | Thr |
| 805  | Ala | Ser | Thr |
| 810  | Ala | Ser | Thr |
| 815  | Ala | Ser | Thr |
| 820  | Ala | Ser | Thr |
| 825  | Ala | Ser | Thr |
| 830  | Ala | Ser | Thr |
| 835  | Ala | Ser | Thr |
| 840  | Ala | Ser | Thr |
| 845  | Ala | Ser | Thr |
| 850  | Ala | Ser | Thr |
| 855  | Ala | Ser | Thr |
| 860  | Ala | Ser | Thr |
| 865  | Ala | Ser | Thr |
| 870  | Ala | Ser | Thr |
| 875  | Ala | Ser | Thr |
| 880  | Ala | Ser | Thr |
| 885  | Ala | Ser | Thr |
| 890  | Ala | Ser | Thr |
| 895  | Ala | Ser | Thr |
| 900  | Ala | Ser | Thr |
| 905  | Ala | Ser | Thr |
| 910  | Ala | Ser | Thr |
| 915  | Ala | Ser | Thr |
| 920  | Ala | Ser | Thr |
| 925  | Ala | Ser | Thr |
| 930  | Ala | Ser | Thr |
| 935  | Ala | Ser | Thr |
| 940  | Ala | Ser | Thr |
| 945  | Ala | Ser | Thr |
| 950  | Ala | Ser | Thr |
| 955  | Ala | Ser | Thr |
| 960  | Ala | Ser | Thr |
| 965  | Ala | Ser | Thr |
| 970  | Ala | Ser | Thr |
| 975  | Ala | Ser | Thr |
| 980  | Ala | Ser | Thr |
| 985  | Ala | Ser | Thr |
| 990  | Ala | Ser | Thr |
| 995  | Ala | Ser | Thr |
| 1000 | Ala | Ser | Thr |

09847513, 0000001

```

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp
145      150
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys
      165      170      175
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
      180      185      190
Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
      195      200      205
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
      210      215      220
Asp Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225      230      235
Asn Val Ala Val Lys Glu Ser Ser Asn Ala
      245      250
<210> 10
<211> 750
<212> DNA
<213> Naturally occurring gamma proteobacterium
<220>
<221> CDS

```

<222> (1)..(750)  
<223> proteorhodopsin variant from clone EBAC41

[illegible]

|   |     |     |     |
|---|-----|-----|-----|
| 115   | 120 | 125 |     |
| cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca     |     |     | 432 |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala     |     |     |     |
| 130   | 135 | 140 |     |
| gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg     |     |     | 480 |
| Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp     |     |     |     |
| 145   | 150 | 155 | 160 |
| gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt     |     |     | 528 |
| Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys     |     |     |     |
| 165   | 170 | 175 |     |
| aat act gca agt cct gct gct gtg caa tca gct tac aac aca atg atg tat |     |     | 576 |
| Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr     |     |     |     |
| 180   | 185 | 190 |     |
| att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt     |     |     | 624 |
| Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly     |     |     |     |
| 195   | 200 | 205 |     |
| tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat     |     |     | 672 |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr             |     |     |     |
| 210   | 215 | 220 |     |
| aac ctt gct gat ttt gtt aac aag att cta ttt ggt tta att ata tgg     |     |     | 720 |
| Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp     |     |     |     |
| 225   | 230 | 235 | 240 |
| aat gtt gct gtt aaa gaa tct tct aat gct                             |     |     | 750 |
| Asn Val Ala Val Lys Glu Ser Ser Asn Ala                             |     |     |     |
| 245   | 250 |     |     |

<210> 11  
 <211> 250  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium  
  
 <400> 11  
  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15  
  
 Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30  
  
 Ser Phe Trp Leu Ala Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110

Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
  
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
 210 215 220  
  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240  
  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250



<210> 12  
 <211> 750  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium  
  
 <220>  
 <221> CDS  
 <222> (1)..(750)  
 <223> Proteorhodopsin variant from clone EBAC64  
  
 <400> 12 48  
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15  
  
 ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt  
 Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30 36  
  
 tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45 144  
  
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60 192  
  
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80 240  
  
 aga gga gta tgg att gaa act ggt gat tcg cct act gta ttt aga tac  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 288

| 85  | 90 | 95 |     |
|---|----|----|-----|
| att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta<br>Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu<br>100 105 110 |    |    | 336 |
| att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt<br>Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu<br>115 120 125     |    |    | 384 |
| cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca<br>Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala<br>130 135 140 |    |    | 432 |
| gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg<br>Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp<br>145 150 155 |    |    | 480 |
| gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt<br>Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys<br>165 170 175 |    |    | 528 |
| aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct<br>Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala<br>180 185 190 |    |    | 576 |
| atc ata gtc ttc ggt tgg gca att tat cct ata ggt tat ttc aca ggt<br>Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Ile Gly Tyr Phe Thr Gly<br>195 200 205 |    |    | 624 |
| tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat<br>Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr<br>210 215 220         |    |    | 672 |
| aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg   |    |    | 720 |

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

750

aat gtt gct gtt aaa gaa tct tct aat gct  
Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 13

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 13

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Lys Ser Ala Cys  
 165 170 175  
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala  
 180 185 190  
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Ile Gly Tyr Phe Thr Gly  
 195 200 205  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 14  
 <211> 750  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium

<220>  
 <221> CDS  
 <222> (1)..(750)  
 <223> Proteorhodopsin variant from pcr clone HOT01m: GenBank# AF349978

<400> 14 48  
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15  
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96  
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30  
 tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc 144  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
 gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
 aga ggg gta tgg att gag acc ggt gat tcg cca act gta ttt aga tac 288  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
 att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta 336  
 Ile Asp Trp Leu Thr Val Pro Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
 att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta 384  
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu  
 115 120 125  
 ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480  
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys  
 165 170 175  
 aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190

```

ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt      624
ile ile ile phe gly trp ala ile tyr pro val gly tyr phe thr gly
195      200

tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat      672
tyr leu met gly asp gly gly ser ala leu asn leu asn leu ile tyr
210      215

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg      720
asn leu ala asp phe val asn lys ile leu phe gly leu ile ile trp
225      230

aat gtt gct gtt aaa gaa tct tct aat gct      750
asn val ala val lys glu ser ser asn ala
245      250

<210> 15
<211> 250
<212> PRT
<213> Naturally occurring gamma proteobacterium

<400> 15

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1      5      10      15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20      25      30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35      40      45

```

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Lys Ala Ala Cys  
 165 170 175  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190



Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 16  
<211> 753  
<212> DNA  
<213> Naturally occurring gamma prtoeobacterium

<220>  
<221> CDS  
<222> (1)..(753)  
<223> Proteorhodopsin variant from pcr clone HOT75m1: GenBank#AF349979

<400> 16 48  
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca  
Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

|   |     |
|---|-----|
| t t t g c t g c t g c t g g t g g c g a t c t a g a t a g t g a t a c t g t t g g t g t t       | 96  |
| P h e A l a A l a A l a G l y A s p L e u A s p I l e S e r A s p T h r V a l G l y V a l       |     |
| 20 25 30  |     |
| t c a t t c t g g c t g t t a c a g c t g g t a t g t t a g c g g c a c t g t g t t c t t t     | 144 |
| S e r P h e T r p L e u V a l T h r A l a G l y M e t L e u A l a T h r V a l P h e P h e       |     |
| 35 40 45  |     |
| t t t g t a g a a g a c c a a g t c a g c g c t a a g t g g a a a c t t c a c t t g c t         | 192 |
| P h e V a l G l u A r g A s p G l n V a l S e r A l a L y s T r p L y s T h r S e r L e u A l a |     |
| 50 55 60  |     |
| g t a t c t g g t t t a a t t a c t g g t a t a g c t t t t g g c a t t a t c t c t a t a t g   | 240 |
| V a l S e r G l y L e u I l e T h r G l y I l e A l a P h e T r p H i s T y r L e u T y r M e t |     |
| 65 70 75 80   |     |
| a g a g g t g t t t g g a t a g a c a c t g g t g a t a c c a c a g t a t t c a g a t a t       | 288 |
| A r g G l y V a l T r p I l e A s p T h r G l y A s p T h r P r o T h r V a l P h e A r g T y r |     |
| 85 90 95  |     |
| a t t g a t t g g t t a t t a a c t g t t c c a t t a c a a t g g t g a g t t c t a t c t a     | 336 |
| I l e A s p T r p L e u L e u T h r V a l P r o L e u G l n M e t V a l G l u P h e T y r L e u |     |
| 100 105 110   |     |
| a t t c t t g c t g c t t g t a c a a g t g t t g c t g c t t c a t t a t t a a g a a g c t t   | 384 |
| I l e L e u A l a A l a C y s T h r S e r V a l A l a A l a S e r L e u P h e L y s L y s L e u |     |
| 115 120 125   |     |
| c t a g c t g g t t c a t t a g t a t g t t a g g t g c t g g a t t t g c a g g c g a a g c t   | 432 |
| L e u A l a G l y S e r L e u V a l M e t L e u G l y A l a G l y P h e A l a G l y G l u A l a |     |
| 130 135 140   |     |
| g g a t t a g c t c c t g t a t t a c c t g c t t t c a t t a t t g g t a t g g c t g g a t g g | 480 |
| G l y L e u A l a P r o V a l L e u P r o A l a P h e I l e I l e G l y M e t A l a G l y T r p |     |
| 145 150 155 160   |     |

|   |     |
|---|-----|
| tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta | 528 |
| Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Lys Ala Ala Val     | 175 |
|   | 170 |
| 165   |     |
| agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg | 576 |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met | 190 |
| 180   |     |
| 185   |     |
| att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt | 624 |
| Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly | 205 |
| 195   |     |
| 200   |     |
| tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata | 672 |
| Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile | 220 |
| 210   |     |
| 215   |     |
| tat aac ctt gcc gac ctt gtt aac aag att cta ttt ggt ttg atc att | 720 |
| Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile | 240 |
| 225   |     |
| 230   |     |
| 235   |     |
| tgg aat gtt gct gtt aaa gaa tct tct aat gct                     | 753 |
| Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala                     | 250 |
| 245   |     |
| <210> 17  |     |
| <211> 251   |     |
| <212> PRT   |     |
| <213> Naturally occurring gamma prtoeobacterium                 |     |
| <400> 17  |     |
| Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser | 15  |
| 1   |     |
| 5   |     |
| 10  |     |

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45  
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Ala  
 50 55 60  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp

145                    150                    155                    160  
  
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
165                    170  
  
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met  
180                    185                    190  
  
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
195                    200                    205  
  
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
210                    215                    220  
  
Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225                    230                    235                    240  
  
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245                    250  
  
<210> 18  
<211> 753  
<212> DNA  
<213> Naturally occurring gamma proteobacterium  
  
<220>  
<221> CDS  
<222> (1) .. (753)

<223> Proteorhodopsin variant from pcr clone HOT75m3; GenBank#AF349980

```
<400> 18
atg ggt aaa tta ctg ata tta ggt agt gct att gca ctt cca tca 48
Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt 144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe
35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg 240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat 288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
```

|   |     |     |     |
|---|-----|-----|-----|
| 115   | 120 | 125 |     |
| cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct     |     |     | 432 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala     |     |     |     |
| 130   | 135 | 140 |     |
| ggt tta gct cct gta tta cct gct tta cct att att ggt att gct gga tgg |     |     | 480 |
| Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp     |     |     |     |
| 145   | 150 | 155 | 160 |
| tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta     |     |     | 528 |
| Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val     |     |     |     |
| 165   | 170 | 175 |     |
| agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg aag     |     |     | 576 |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys     |     |     |     |
| 180   | 185 | 190 |     |
| att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt     |     |     | 624 |
| Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly     |     |     |     |
| 195   | 200 | 205 |     |
| tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata     |     |     | 672 |
| Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile     |     |     |     |
| 210   | 215 | 220 |     |
| tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att     |     |     | 720 |
| Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile     |     |     |     |
| 225   | 230 | 235 | 240 |
| tgg aat gtt gct gtt aaa gaa tct tct aat gct                         |     |     | 753 |
| Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala                         |     |     |     |
| 245   | 250 |     |     |

<210> 19

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 19

Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
100 105 110



Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
  
 Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys  
 180 185 190  
  
 Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
 195 200 205  
  
 Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
 210 215 220  
  
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
 225 230 235 240  
  
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 20  
<211> 753  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(753)  
<223> Proteorhodopsin variant from pcr clone HOT75m4; GenBank #AF349981

[illegible]

09047513 030303

| 85  | 90      | 95  |     |
|---|---------|-----|-----|
| att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta |         |     | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Glu Phe Tyr Leu     | 100 105 | 110 |     |
| att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt |         |     | 384 |
| Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu | 115 120 | 125 |     |
| cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct |         |     | 432 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala | 130 135 | 140 |     |
| gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg |         |     | 480 |
| Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp | 145 150 | 155 |     |
| tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta |         |     | 528 |
| Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val | 165 170 | 175 |     |
| agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg |         |     | 576 |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met     | 180 185 | 190 |     |
| att att gtt gga tgg gca att tat cct gct gga tat gct gct ggt     |         |     | 624 |
| Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly | 195 200 | 205 |     |
| tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata |         |     | 672 |
| Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Ile         | 210 215 | 220 |     |
| tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att |         |     | 720 |

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 21

<211> 251

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 21

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Tyr Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met  
 180 185 190  
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
 195 200 205  
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 22  
<211> 753  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(753)  
<223> Proteorhodopsin variant from pcr clone HOT75m8: GenBank#AF349982

<400> 22 48  
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca  
Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

ttt gct gct gct ggc gat cta gat ata agt gat act gtt ggt gtt 96  
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144  
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192

|   |     |
|---|-----|
| Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr         |     |
| 50  | 60  |
| 55  |     |
| gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg         | 240 |
| Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met         |     |
| 65  | 75  |
| 70  | 80  |
| aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat         | 288 |
| Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr         |     |
| 85  | 95  |
| 90  |     |
| att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta         | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu         |     |
| 100   | 110 |
| 105   |     |
| att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt         | 384 |
| Ile Leu Ala Ala Cys Thr Asn Val Asn Val Ala Ala Ser Leu Phe Lys Lys Leu |     |
| 115   | 125 |
| 120   |     |
| cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct         | 432 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala         |     |
| 130   | 140 |
| 135   |     |
| gga ttg gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg         | 480 |
| Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp         |     |
| 145   | 155 |
| 150   | 160 |
| tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta         | 528 |
| Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val         |     |
| 165   | 175 |
| 170   |     |
| agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg gtg         | 576 |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val         |     |
| 180   | 185 |
| 185   | 190 |

att att gtt gtt gga tgg gca att tat cct gct gct gct gct ggt 624  
 ile ile val val gly trp ala ile tyr pro ala gly tyr ala ala gly  
 195 200 205

tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672  
 tyr leu met gly gly glu gly val tyr ala ser asn leu asn leu ile  
 210 215 220

tat aac ctt gcc gac ctt gtt aac aag att cta ttt ggt ttg atc att 720  
 tyr asn leu ala asp leu val asn lys ile leu phe gly leu ile ile  
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753  
 trp asn val ala val lys glu ser ser asn ala  
 245 250

<210> 23  
 <211> 251  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 23  
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45

09847513.080801



Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
 Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val  
 180 185 190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
210 215 220

Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 24  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB0m1: GenBank#AF349983

<400> 24 48  
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15  
ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96

|   |     |
|---|-----|
| Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val |     |
| 20 25 30  |     |
| tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc | 144 |
| Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe |     |
| 35 40 45  |     |
| ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act | 192 |
| Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |     |
| 50 55 60  |     |
| gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg | 240 |
| Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met |     |
| 65 70 75 80   |     |
| aga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac | 288 |
| Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr |     |
| 85 90 95  |     |
| att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu |     |
| 100 105 110   |     |
| att ctt gct gct gca aca aat gtt gct gct ggc ctg ttt aag aaa tta | 384 |
| Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu     |     |
| 115 120 125   |     |
| ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca | 432 |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala |     |
| 130 135 140   |     |
| gga att atg aac gct tgg cct gca ttc att att ggg tgt tta gct tgg | 480 |
| Gly Ile Met Asn Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp |     |
| 145 150 155 160   |     |

gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys 175  
 165 170  
 aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct 576  
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala 190  
 180 185  
 atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly 205  
 195 200  
 tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat 672  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr 220  
 210 215  
 aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp 235  
 225 230  
 aat gtt gct gtt aaa gaa tct tct aat gct 750  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250  
 245  
 <210> 25  
 <211> 250  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium  
 <400> 25  
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr 15  
 1 5 10

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30  
  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 Gly Ile Met Asn Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys  
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala  
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 26  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB0m2

```

<400> 26
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu 5 Ile Leu Gly Ser Val 10 Ile Ala Leu Pro Thr 15
1
ttt gct gca ggt ggt ggt gac gct gct ggt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser 25 Tyr Thr Gly Val 30
20
tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser 40 Thr Val Phe Phe 45
35
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50
55
gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe 70 Trp His Tyr Met Tyr Met 75
65
aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp 85 Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr 90
95
att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu 100
105
att ctt gct gct gct act aat gtt gct gct ggc ctg ttt aag aaa tta 384
Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu 110
115

```

ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480  
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 gta tac atg att tat gag ctt tgg ctt gga gaa gga aaa gct gcg tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Trp Leu Gly Glu Gly Lys Ala Ala Cys  
 165 170 175  
 aat aca gca agt cct gct gtt cag tca gct tac aac aca atg atg atg 576  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Met  
 180 185 190  
 atc atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
 tac cta atg ggt gac ggt gga tca gca ctt aac tta aac ctt atc tat 672  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
 210 215 220  
 aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240  
 aat gtt gct gtt aaa gaa tct tct aat gct 750  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250  
 <210> 27  
 <211> 250



<212> PRT  
 <213> Naturally occurring gamma proteobacterium  
  
 <400> 27  
  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15  
  
 Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30  
  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu  
 115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
130 135 140

Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp  
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Trp Leu Gly Glu Lys Ala Ala Cys  
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Met  
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 28

<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB20m2; GenBank #AF349985

[illegible]

att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
  
 att ctt gct gct gca act aat gtt gct gct ggc ctg ttt aag aaa tta 384  
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu  
 115 120 125  
  
 ttg gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gag gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg 480  
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ala Ala Cys  
 165 170 175  
  
 aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
  
 ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
  
 tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672  
 Tyr Leu Met Gly Asp Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr  
 210 215 220  
  
 aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp

|   |     |     |     |     |
|---|-----|-----|-----|-----|
| 225   | 230 | 235 | 240 | 750 |
| aat gtt gct gtt aaa gaa tct tct aat gct                         |     |     |     |     |
| Asn Val Ala Val Lys Glu Ser Ser Asn Ala                         |     |     |     |     |
|   | 245 | 250 |     |     |
| <210> 29  |     |     |     |     |
| <211> 250   |     |     |     |     |
| <212> PRT   |     |     |     |     |
| <213> Naturally occurring gamma proteobacterium                 |     |     |     |     |
| <400> 29  |     |     |     |     |
| Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr |     |     |     |     |
| 1   | 5   | 10  | 15  |     |
| Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val |     |     |     |     |
|   | 20  | 25  | 30  |     |
| Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe |     |     |     |     |
|   | 35  | 40  | 45  |     |
| Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |     |     |     |     |
|   | 50  | 55  | 60  |     |
| Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met |     |     |     |     |
| 65  | 70  | 75  | 80  |     |
| Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr |     |     |     |     |

09847513.0000001

|   |     |     |     |
|---|-----|-----|-----|
| Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu | 85  | 90  | 95  |
| 100   | 105 | 110 |     |
| Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu     | 115 | 120 | 125 |
|   |     |     |     |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala | 130 | 135 | 140 |
|   |     |     |     |
| Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp | 145 | 150 | 155 |
|   |     |     | 160 |
| Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Lys Ala Ala Cys     | 165 | 170 | 175 |
|   |     |     |     |
| Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr | 180 | 185 | 190 |
|   |     |     |     |
| Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly | 195 | 200 | 205 |
|   |     |     |     |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr         | 210 | 215 | 220 |
|   |     |     |     |

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 30  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB20m5; GenBank#AF349986

<400> 30 48  
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96  
Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144  
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192  
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
 aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
 att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta 336  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
 att ctt gct gct act aat gtt gct gga tca tta ttt aag aaa tta 384  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
 cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 caa att atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480  
 Gln Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
 aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct 576  
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala  
 180 185 190  
 atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly



|   |     |     |     |
|---|-----|-----|-----|
| 195   | 200 | 205 |     |
| tac cta atg ggt gac ggt ggg tca gct ctt aac tta aac ctt att tat |     |     | 672 |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr         |     |     |     |
| 210   | 215 | 220 |     |
| aac ctt gct gac ttt gtt aac aag att cta ctt ggt tta att ata tgg |     |     | 720 |
| Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Gly Leu Ile Trp         |     |     |     |
| 225   | 230 | 235 | 240 |
| aat gtt gct gtt aaa gaa tct tct aat gct                         |     |     | 750 |
| Asn Val Ala Val Lys Glu Ser Ser Asn Ala                         |     |     |     |
| 245   | 250 |     |     |

<210> 31  
 <211> 250  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 31

|   |    |
|---|----|
| Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr |    |
| 1   | 15 |

|   |    |
|---|----|
| Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val |    |
| 20  | 30 |

|   |    |
|---|----|
| Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe |    |
| 35  | 45 |

|   |  |
|---|--|
| Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |  |
|---|--|

50

55

60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
100 105 110

Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
130 135 140

Gln Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
145 150 155 160

Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Lys Ser Ala Cys  
165 170 175

Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala  
180 185 190

Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Leu Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 32  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB20ml2; GenBank #AF349987

<400> 32 48  
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96  
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

|   |     |
|---|-----|
| tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc | 144 |
| Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe |     |
| 35 40 45  |     |
| ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act | 192 |
| Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |     |
| 50 55 60  |     |
| gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg | 240 |
| Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met |     |
| 65 70 75 80   |     |
| aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac | 288 |
| Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr |     |
| 85 90 95  |     |
| att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu |     |
| 100 105 110   |     |
| att ctt gct gca gct aat gtt gct gga tca tta ttt aag aaa tta     | 384 |
| Ile Leu Ala Ala Ala Ala Asn Val Ala Gly Ser Leu Phe Lys Lys Leu |     |
| 115 120 125   |     |
| cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca | 432 |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala |     |
| 130 135 140   |     |
| gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg | 480 |
| Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp |     |
| 145 150 155 160   |     |
| gta tac atg att tat gaa tta tgg gct gga gaa gga aaa tct gca tgt | 528 |
| Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys |     |

|   |     |     |     |
|---|-----|-----|-----|
| 165   | 170 | 175 |     |
| aat act gca agt cct gct gtg caa tca gcc tac aac aca atg atg tat |     |     | 576 |
| Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr |     | 190 |     |
| 180   | 185 |     |     |
| att atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt |     |     | 624 |
| Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly |     | 205 |     |
| 195   | 200 |     |     |
| tac ttg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat |     |     | 672 |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr         |     | 220 |     |
| 210   | 215 |     |     |
| aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg |     |     | 720 |
| Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp |     | 235 |     |
| 225   | 230 |     |     |
| aat gtt gct gtt aaa gaa tct tct aat gct                         |     |     | 750 |
| Asn Val Ala Val Lys Glu Ser Ser Asn Ala                         |     | 250 |     |
| 245   |     |     |     |

<210> 33  
 <211> 250  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 33

|   |    |
|---|----|
| Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr |    |
| 1   | 5  |
|   | 10 |
|   | 15 |

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val

|   |     |     |     |
|---|-----|-----|-----|
|   | 20  | 25  | 30  |
| Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe |     |     |     |
| 35  | 40  | 45  |     |
| Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |     |     |     |
| 50  | 55  | 60  |     |
| Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met |     |     |     |
| 65  | 70  | 75  | 80  |
| Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr |     |     |     |
| 85  | 90  | 95  |     |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu |     |     |     |
| 100   | 105 | 110 |     |
| Ile Leu Ala Ala Ala Asn Val Ala Gly Ser Leu Phe Lys Lys Leu     |     |     |     |
| 115   | 120 | 125 |     |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala |     |     |     |
| 130   | 135 | 140 |     |
| Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp |     |     |     |
| 145   | 150 | 155 | 160 |

Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Lys Ser Ala Cys  
165 170 175

Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
180 185 190

Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 34  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB40ml; GenBank #AF349988

<400> 34  
 atg ggt aaa tta tta ctg ata ata ggt agt gtt att gca ctt cct aca 48  
 Met Gly Lys Leu Leu 5 Ile Ile Gly Ser Val 10 Ile Ala Leu Pro Thr 15  
 1  
 ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96  
 Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser 25 Thr Tyr Gly Val 30  
 20  
 tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser 40 Thr Val Phe Phe 45  
 35  
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr 50  
 55  
 gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met 70  
 65  
 aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr 85  
 90  
 att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu 100  
 105  
 att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt 384  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu 115  
 120  
 cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432



```

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130                               135       140
gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145                               150       155       160
gta tat atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt
Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165                               170       175
aat aca gca agt cct gct gct gtg caa tca gct tac aac aca atg atg tat
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr
180                               185       190
att atc gtc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt
Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195                               200       205
tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
210                               215       220
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225                               230       235       240
aat gtt gct gtt aaa gaa tct tct aat gct
Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245                               250

```

<210> 35  
 <211> 250  
 <212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 35

Met Gly Lys Leu Leu Ile Ile Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
  
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
 210 215 220  
  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240  
  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 36  
 <211> 750

<212> DNA  
 <213> Naturally occurring gamma proteobacterium  
  
 <220>  
 <221> CDS  
 <222> (1)..(750)  
 <223> Proteorhodopsin variant from pcr clone MB40m5;p GenBank #AF349989  
  
 <400> 36 48  
 atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15  
  
 ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96  
 Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30  
  
 tct ttt tgg tta gtt act gct gct cta tta gca tct act gta ttt ttc 144  
 Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
  
 ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act 192  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 gta tcg ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg 240  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
  
 aga ggg gta tgg att gag act ggt gat tcg cca act gta ttt aga tac 288  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 att gat tgg tta cta aca gtt cct cta ttg ata tgt gaa ttc tac tta 336

|   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile Asp Trp   | Leu | 100 | Thr | Val | Pro | Leu | 105 | Ile | Cys | Glu | Phe | Tyr | Leu |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 110 |
| att ctt gct gct gca aca aat gtt gct gct gct ggc ctg ttt aag aaa tta |     |     |     |     |     |     |     |     |     |     |     |     |     | 384 |
| Ile Leu Ala Ala Thr Asn Val 120                                     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 125 |
| ttg gtt ggt tct ctt gtt atg ctt gtt gtt ggt tac atg ggt gag gca     |     |     |     |     |     |     |     |     |     |     |     |     |     | 432 |
| Leu Val Gly Ser Leu Val Met 135                                     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 140 |
| gga att atg aac gct tgg ggt gca ttc gtt att ggg tgt tta gct tgg     |     |     |     |     |     |     |     |     |     |     |     |     |     | 480 |
| Gly Ile Met Asn Ala Trp Gly Ala Phe Val 150                         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 155 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 160 |
| gta tac atg att tat gaa cta tgg gct gga gaa ggc aag gct gca tgt     |     |     |     |     |     |     |     |     |     |     |     |     |     | 528 |
| Val Tyr Met Ile Tyr 165   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 170 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 175 |
| aat act gca agt cct gct gct gtt gca tca gct tac aac aca atg atg tat |     |     |     |     |     |     |     |     |     |     |     |     |     | 576 |
| Asn Thr Ala Ser Pro Ala Val 180                                     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 185 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 190 |
| ata atc atc ttt ggt tgg gca att tat cct gta ggt tat ttc aca ggt     |     |     |     |     |     |     |     |     |     |     |     |     |     | 624 |
| Ile Ile Ile Phe Gly Trp Ala 195                                     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 200 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 205 |
| tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat     |     |     |     |     |     |     |     |     |     |     |     |     |     | 672 |
| Tyr Leu Met Gly Asp Gly 210   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 215 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 220 |
| aac ctt gct gac ttt gtt aac aag aat cta ttt ggt tta att ata tgg     |     |     |     |     |     |     |     |     |     |     |     |     |     | 720 |
| Asn Leu Ala Asp Phe Val 225   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 230 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 235 |
|   |     |     |     |     |     |     |     |     |     |     |     |     |     | 240 |

750

aat gtt gct gtt aaa gaa tct tct aat gct  
Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 37

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 37

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Thr Asn Val Ala Ala Gly Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 Gly Ile Met Asn Ala Trp Gly Ala Phe Val Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Lys Ala Ala Cys  
 165 170 175  
  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
  
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
 210 215 220  
  
 Asn Leu Ala Asp Phe Val Asn Lys Asn Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 38  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB40ml2; GenBank # AF34999

<400> 38 48  
atg ggt aaa tta tta cgg ata tta ggt agt gtt att gca ctt cct aca  
Met Gly Lys Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15  
ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96  
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30  
tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc 144  
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45  
ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act 192  
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60



gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
 aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
 att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
 att ctt gct gct gca act aat gtt gct gga tca tta ttt aag aaa tta 384  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
 cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
 aat act gca agt cct gct gtg cca tca gct tac aac aca atg atg tat 576  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
 atc atc atc gtt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205

tac ctg atg ggt gac ggt gga tca gct ctt aac tta aac ctt atc tat 672  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Ile Tyr  
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240

aat gtt gct gtt aaa gaa tct tct aat gct 750  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 39  
 <211> 250  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 39

Met Gly Lys Leu Leu Arg Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
 1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
 20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
 Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly

195

200

205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 40

<211> 750

<212> DNA

<213> Naturally occurring gamma proteobacterium

<220>

<221> CDS

<222> (1)..(750)

<223> Proteorhodopsin variant from pcr clone MB100m5; GenBank #AF349991

<400> 40

atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48  
Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt 96  
Phe Ala Ala Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

```

tct ttt tgg tta gtt aca gct gct cta tta gca tct act gta ttt ttc      144
Ser Phe Trp Leu Val Thr Ala Thr Ala Leu Leu Ala Ser Thr Val Phe Phe
35                                40      45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act      192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50                                55      60

gta tct ggt ctt gtt act ggt att gct ttc ttc tgg cat tac atg tac atg      240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65                                70      75      80

aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac      288
Arg Gly Val Trp Ile Glu Thr Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85                                90      95

att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta      336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100                               105      110

att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt      384
Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu
115                               120      125

cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca      432
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala
130                               135      140

gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg      480
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp
145                               150      155      160

gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt      528
Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys
165                               170      175

```

```

aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct      576
Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala
180                                     185      190

atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt      624
Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly
195                                     200      205

tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat      672
Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Leu Ile Tyr
210                                     215      220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg      720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225                                     230      235      240

aat gtt gct gtt aaa gaa gaa tct tct aat gct      750
Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245                                     250

<210> 41
<211> 250
<212> PRT
<213> Naturally occurring gamma proteobacterium

<400> 41

Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1      5      10      15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20      25      30

```

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Lys Ser Ala Cys

|   |     |     |
|---|-----|-----|
| 165   | 170 | 175 |
| Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala         |     |     |
| 180   | 185 | 190 |
| Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly         |     |     |
| 195   | 200 | 205 |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr                 |     |     |
| 210   | 215 | 220 |
| Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp         |     |     |
| 225   | 230 | 235 |
| Asn Val Ala Val Lys Glu Ser Ser Asn Ala                                 |     |     |
|   | 245 | 250 |
| <210> 42  |     |     |
| <211> 750   |     |     |
| <212> DNA   |     |     |
| <213> Naturally occurring gamma proteobacterium                         |     |     |
| <220>   |     |     |
| <221> CDS   |     |     |
| <222> (1) .. (750)  |     |     |
| <223> Proteorhodopsin variant from pcr clone MB100m7; GenBank #AF349992 |     |     |
| <400> 42  |     |     |



|   |     |
|---|-----|
| atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca<br>Met Gly Lys Leu Leu 5<br>1   | 48  |
| ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt<br>Phe Ala Ala Gly Gly 20<br>25   | 96  |
| tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc<br>Ser Phe Trp Leu Val Thr Ala 40<br>35 45  | 144 |
| ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tca tta act<br>Phe Val Glu Arg Asp Arg Val 55<br>50 60  | 192 |
| gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tac atg<br>Val Ser Gly Leu Val Thr Gly 70<br>65 75 80                                     | 240 |
| aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac<br>Arg Gly Val Trp Ile Glu Thr Gly Asp Ser 90<br>85 95                            | 288 |
| att gat tgg tta cta aca gtt cct cta tta ata tgt gaa ttc tac tta<br>Ile Asp Trp Leu Leu Thr Val Pro Leu Leu 105<br>100 110                         | 336 |
| att ctt gct gct gct act aat gtt gcc ggc tca tta ttt aag aaa ctt<br>Ile Leu Ala Ala 115<br>120   | 384 |
| cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca<br>Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala<br>130 135 140 | 432 |

gga att atg gca gct tgg cct gca ttc att att ggg tgt tta gct tgg 480  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160

gta tac atg att tat gaa cta tat gct gga gaa gga aaa tct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175

aat act gca agt cct tcg gtt caa tca gct tac aac aca atg atg gct 576  
 Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala  
 180 185 190

atc ata gtc ttc ggt tgg gca att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205

tac cta atg ggt gac ggt gga tca gct ctt aac tta aac ctt att tat 672  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Ile Tyr  
 210 215 220

aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240

aat gct gct gtt aaa gaa tct tct aat gct 750  
 Asn Ala Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 43

<211> 250

<212> PRT

<213> Naturally occurring gamma proteobacterium

<400> 43

Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
100 105 110

Ile Leu Ala Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
115 120 125

Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala

|   |     |     |     |
|---|-----|-----|-----|
| 130   | 135 | 140 |     |
| Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp |     |     |     |
| 145   | 150 | 155 | 160 |
| Val Tyr Met Ile Tyr Glu Leu Tyr Ala Gly Glu Gly Lys Ser Ala Cys |     |     |     |
| 165   | 170 |     | 175 |
| Asn Thr Ala Ser Pro Ser Val Gln Ser Ala Tyr Asn Thr Met Met Ala |     |     |     |
| 180   | 185 |     | 190 |
| Ile Ile Val Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly |     |     |     |
| 195   | 200 | 205 |     |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr         |     |     |     |
| 210   | 215 | 220 |     |
| Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp |     |     |     |
| 225   | 230 | 235 | 240 |
| Asn Ala Ala Val Lys Glu Ser Ser Asn Ala                         |     |     |     |
|   | 245 | 250 |     |

<210> 44  
 <211> 750  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium

<220>  
 <221> CDS  
 <222> (1)..(750)  
 <223> Proteorhodopsin variant from pcr clone MB100m9; GenBank #AF349993

```

<400> 44
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca 48
Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

ttt gct gca ggt ggt ggt gac ctt gat gct agt gat tac act ggt gtt 96
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

tct ttt tgg tta gtt act gct gct tta tta gca tct act gta ttt ttc 144
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act 192
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg 240
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met
65 70 75 80

aga ggg gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac 288
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr
85 90 95

ata gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu
100 105 110

```

att ctt gcc gct gca act aat gtt gct gga tca tta ttt aag aaa tta 384  
 Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 ctt gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca 432  
 Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
 130 135 140  
  
 gga atc atg gct gca tgg cct gca ttc att att ggg tgt tta gct tgg 480  
 Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp  
 145 150 155 160  
  
 gta tac atg att tat gaa cta tgg gct gga gaa gga aaa tct gca tgt 528  
 Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys  
 165 170 175  
  
 aat act gca agt cct gct gtg caa tca gct tac aac aca atg atg tat 576  
 Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
 180 185 190  
  
 atc atc atc ttt ggt tgg gcg att tat cct gta ggt tat ttc aca ggt 624  
 Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
 195 200 205  
  
 tac ctt atg ggt gac ggt gga tca gca ctt aac tta aac ctt att tat 672  
 Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Asn Ile Tyr  
 210 215 220  
  
 aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720  
 Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp  
 225 230 235 240  
  
 aat gtt gct gtt aaa gaa tct tct aat gct 750  
 Asn Val Ala Val Lys Glu Ser Ser Asn Ala



|   |     |     |     |
|---|-----|-----|-----|
| Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu     | 100 | 105 | 110 |
| 115   | 120 | 125 |     |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala | 130 | 135 | 140 |
| 145   | 150 | 155 | 160 |
| Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Ile Gly Cys Leu Ala Trp | 165 | 170 | 175 |
| 180   | 185 | 190 |     |
| Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr | 195 | 200 | 205 |
| Ile Ile Ile Phe Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly | 210 | 215 | 220 |
| Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr         | 225 | 230 | 235 |
| Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp | 240 |     |     |



Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 46  
<211> 750  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(750)  
<223> Proteorhodopsin variant from pcr clone MB100m10; GenBank #AF34999

<400> 46 48  
atg ggt aaa tta tta ctg ata tta ggt agt gtt att gca ctt cct aca  
Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr  
1 5 10 15

ttt gct gca ggt ggc ggt gac ctt gat gct agt gat tac act ggt gtt  
Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val  
20 25 30 96

tct ttt tgg tta gtt aca gct gct cta tta gcg tct act gta ttt ttc  
Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe  
35 40 45 144

ttt gtt gaa aga gat aga gtt tct gca aaa tgg aaa aca tta act  
Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60 192

gta tct ggt ctt gtt act ggt att gct ttc tgg cat tac atg tat atg  
Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
240

|   |  |     |    |     |  |     |     |
|---|--|-----|----|-----|--|-----|-----|
| 65  |  | 70  |    | 75  |  | 80  |     |
| aga gga gta tgg att gaa act ggt gat tcg cca act gta ttt aga tac |  |     |    |     |  |     | 288 |
| Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr |  |     |    |     |  |     |     |
| 85  |  |     | 90 |     |  | 95  |     |
| att gat tgg tta cta aca gtt cct tta tta ata tgt gaa ttc tac tta |  |     |    |     |  |     | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Ile Cys Glu Phe Tyr Leu     |  |     |    |     |  |     |     |
| 100   |  | 105 |    | 110 |  |     |     |
| att ctt gct gct gca act aat gtt gcc ggc tca tta ttt aag aaa ctt |  |     |    |     |  |     | 384 |
| Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu     |  |     |    |     |  |     |     |
| 115   |  | 120 |    | 125 |  |     |     |
| cta gtt ggt tct ctt gtt atg ctt gtg ttt ggt tac atg ggt gaa gca |  |     |    |     |  |     | 432 |
| Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala |  |     |    |     |  |     |     |
| 130   |  | 135 |    | 140 |  |     |     |
| gga ata atg gcg gct tgg cct gca ttc atc gtt gga tgt tta gca tgg |  |     |    |     |  |     | 480 |
| Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Val Gly Cys Leu Ala Trp |  |     |    |     |  |     |     |
| 145   |  | 150 |    | 155 |  | 160 |     |
| gta tat atg att tat gaa cta tgg gct ggt gaa gga aaa tct gca tgt |  |     |    |     |  |     | 528 |
| Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Lys Ser Ala Cys     |  |     |    |     |  |     |     |
| 165   |  | 170 |    | 175 |  |     |     |
| aat act gca agt cct gct gta cag tca gct tac aac aca atg atg tat |  |     |    |     |  |     | 576 |
| Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr |  |     |    |     |  |     |     |
| 180   |  | 185 |    | 190 |  |     |     |
| atc atc atc gtt ggt tgg gca att tat cct gta ggt tat ttc aca ggt |  |     |    |     |  |     | 624 |
| Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly |  |     |    |     |  |     |     |
| 195   |  | 200 |    | 205 |  |     |     |
| tac cta atg ggt gac ggt gga tca gct ctt aat cta aac ctt att tat |  |     |    |     |  |     | 672 |

```

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr
210 215 220
aac ctt gct gac ttt gtt aac aag att cta ttt ggt tta att ata tgg 720
Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile Trp
225 230 235 240
aat gtt gct gtt aaa gaa tct tct aat gct
Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250
245

<210> 47
<211> 250
<212> PRT
<213> Naturally occurring gamma proteobacterium

<400> 47

Met Gly Lys Leu Leu Ile Leu Gly Ser Val Ile Ala Leu Pro Thr
1 5 10 15

Phe Ala Ala Gly Gly Gly Asp Leu Asp Ala Ser Asp Tyr Thr Gly Val
20 25 30

Ser Phe Trp Leu Val Thr Ala Ala Leu Leu Ala Ser Thr Val Phe Phe
35 40 45

Phe Val Glu Arg Asp Arg Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

```

Val Ser Gly Leu Val Thr Gly Ile Ala Phe Trp His Tyr Met Tyr Met  
65 70 75 80  
Arg Gly Val Trp Ile Glu Thr Gly Asp Ser Pro Thr Val Phe Arg Tyr  
85 90 95  
Ile Asp Trp Leu Leu Thr Val Pro Leu Leu Ile Cys Glu Phe Tyr Leu  
100 105 110  
Ile Leu Ala Ala Thr Asn Val Ala Gly Ser Leu Phe Lys Lys Leu  
115 120 125  
Leu Val Gly Ser Leu Val Met Leu Val Phe Gly Tyr Met Gly Glu Ala  
130 135 140  
Gly Ile Met Ala Ala Trp Pro Ala Phe Ile Val Gly Cys Leu Ala Trp  
145 150 155 160  
Val Tyr Met Ile Tyr Glu Leu Trp Ala Gly Glu Gly Lys Ser Ala Cys  
165 170 175  
Asn Thr Ala Ser Pro Ala Val Gln Ser Ala Tyr Asn Thr Met Met Tyr  
180 185 190  
Ile Ile Ile Val Gly Trp Ala Ile Tyr Pro Val Gly Tyr Phe Thr Gly  
195 200 205

Tyr Leu Met Gly Asp Gly Gly Ser Ala Leu Asn Leu Ile Tyr  
 210 215 220

Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Trp  
 225 230 235 240

Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 48  
 <211> 753  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium

<220>  
 <221> CDS  
 <222> (1)..(753)  
 <223> Proteorhodopsin variant from pcr clone PALB1; GenBank #AF349995

<400> 48  
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48  
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96  
 Phe Ala Ala Ala Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt 144  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe

|   |     |     |     |
|---|-----|-----|-----|
| 35  | 40  | 45  |     |
| ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act |     |     | 192 |
| Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |     |     |     |
| 50  | 55  | 60  |     |
| gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg |     |     | 240 |
| Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met |     |     |     |
| 65  | 70  | 75  | 80  |
| aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat |     |     | 288 |
| Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr |     |     |     |
| 85  | 90  | 95  |     |
| att gat tgg cta tta act gtt cca tta caa atg gtt gag ttc tat cta |     |     | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu |     |     |     |
| 100   | 105 | 110 |     |
| att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt |     |     | 384 |
| Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Leu     |     |     |     |
| 115   | 120 | 125 |     |
| cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct |     |     | 432 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala |     |     |     |
| 130   | 135 | 140 |     |
| ggt tta gct cct gta tta cct gct ttc att ctt ggt atg gct ggt tgg |     |     | 480 |
| Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Leu Gly Met Ala Gly Trp |     |     |     |
| 145   | 150 | 155 | 160 |
| tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta |     |     | 528 |
| Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val |     |     |     |
| 165   | 170 | 175 |     |
| agt act gca agt cct gct gtt aac tct gct tac aat gca atg atg aag |     |     | 576 |

```

Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys
180                                     185          190
att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt
Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195                                     200          205          624

tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata
Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210                                     215          220          672

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att
Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225                                     230          235          720

tgg aat gtt gct gtt aaa gaa tct tct aat gct
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245                                     250          753

<210> 49
<211> 251
<212> PRT
<213> Naturally occurring gamma proteobacterium

<400> 49

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1      5      10      15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20     25     30

```

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Leu Gly Met Ala Gly Trp  
 145 150 155 160  
  
 Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175



Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys  
180 185 190

Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
195 200 205

Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 50  
<211> 753  
<212> DNA  
<213> Naturally occurring gamma proteobacterium

<220>  
<221> CDS  
<222> (1)..(753)  
<223> Proteorhodopsin variant from pcr clone PALB2; GenBank #AF349996

<400> 50 48  
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca  
Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser

|   |     |     |     |     |
|---|-----|-----|-----|-----|
| 1   | 5   | 10  | 15  |     |
| ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt |     |     |     | 96  |
| Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val     | 20  | 25  | 30  |     |
| tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt | 35  | 40  | 45  | 144 |
| Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe     |     |     |     |     |
| ttt gta gaa aga gac caa gtc agc gct gag tgg aaa act tca ctt act |     |     |     | 192 |
| Phe Val Glu Arg Asp Gln Val Ser Ala Glu Trp Lys Thr Ser Leu Thr | 50  | 55  | 60  |     |
| gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg |     |     |     | 240 |
| Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met | 65  | 70  | 75  |     |
| aga ggt gtt tgg ata gat act ggt gat acc cca aca gta ttc aga tat |     |     |     | 288 |
| Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr | 85  | 90  | 95  |     |
| att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta |     |     |     | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu | 100 | 105 | 110 |     |
| att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt |     |     |     | 384 |
| Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu | 115 | 120 | 125 |     |
| cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct |     |     |     | 432 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala | 130 | 135 | 140 |     |
| gga tta gct cct gta tta cct gct gct ttc att att ggt atg gga tgg |     |     |     | 480 |

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val 175  
 165 170  
 agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met 190  
 185  
 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624  
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly 205  
 195 200  
 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672  
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile 220  
 215  
 tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720  
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile 240  
 225 230  
 tgg aat gtt gct gtt aaa gaa tct tct aat gct 753  
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala 250  
 245  
 <210> 51  
 <211> 251  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium  
 <400> 51

Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15  
  
 Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30  
  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe  
 35 40 45  
  
 Phe Val Glu Arg Asp Gln Val Ser Ala Glu Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140

Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155  
  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met  
 180 185 190  
  
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
 195 200 205  
  
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
 210 215 220  
  
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
 225 230 235 240  
  
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250  
  
 <210> 52  
 <211> 753  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium  
  
 <220>

<221> CDS  
 <222> (1)..(753)  
 <223> Proteorhodopsin variant from pcr clone PALB5; GenBank#AF349997

```

<400> 52
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
Phe Ala Ala Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe
35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt tta att act ggt ata gcc ttt tgg cat tat ctc tat atg 240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu
100 105 110

att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt 384

```

|     |     |     |     |     |     |     |      |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Leu | Ala | Ala | Cys | Thr | Asn | Val  | Ala | Ala | Ser | Leu | Phe | Lys | Lys | Leu |     |
|     |     | 115 |     |     |     |     | 120  |     |     |     |     | 125 |     |     |     |     |
| cta | gct | ggt | tca | tta | gta | atg | tta  | ggt | gct | gga | ttt | gca | ggc | gaa | gct | 432 |
| Leu | Ala | Gly | Ser | Leu | Val | Met | Leu  | Gly | Ala | Gly | Phe | Ala | Gly | Glu | Ala |     |
|     |     | 130 |     |     |     | 135 |      |     |     |     | 140 |     |     |     |     |     |
| gga | tta | gct | cct | gta | tgg | cct | gct  | ttc | att | att | ggt | atg | gct | gga | tgg | 480 |
| Gly | Leu | Ala | Pro | Val | Trp | Pro | Ala  | Phe | Ile | Ile | Gly | Met | Ala | Gly | Trp |     |
| 145 |     |     |     |     | 150 |     |      |     | 155 |     |     |     |     |     | 160 |     |
| tta | tac | atg | att | tat | gag | cta | tat  | atg | ggt | gaa | ggt | aag | gct | gct | gta | 528 |
| Leu | Tyr | Met | Ile | Tyr | Glu | Leu | Tyr  | Met | Gly | Glu | Gly | Lys | Ala | Ala | Val |     |
|     |     |     | 165 |     |     |     |      |     | 170 |     |     |     |     |     | 175 |     |
| agt | act | gca | agt | cct | gct | gct | ggt  | aac | tct | gca | tac | aac | gca | atg | atg | 576 |
| Ser | Thr | Ala | Ser | Pro | Ala | Val | Asn  | Ser | Ala | Tyr | Asn | Ala | Met | Met | Met |     |
|     |     |     | 180 |     |     |     |      | 185 |     |     |     |     |     |     |     |     |
| att | att | ggt | ggt | gga | tgg | gca | att  | tat | cct | gct | gga | tat | gct | gct | ggt | 624 |
| Ile | Ile | Val | Val | Gly | Trp | Ala | Ile  | Tyr | Pro | Ala | Gly | Tyr | Ala | Ala | Gly |     |
|     |     |     | 195 |     |     |     | 200. |     |     |     |     | 205 |     |     |     |     |
| tac | cta | atg | ggt | ggc | gaa | ggt | gta  | tac | gct | tca | aac | cta | aac | ctt | ata | 672 |
| Tyr | Leu | Met | Gly | Gly | Glu | Glu | Gly  | Val | Tyr | Ala | Ser | Asn | Leu | Asn | Ile |     |
|     |     | 210 |     |     |     | 215 |      |     |     |     | 220 |     |     |     |     |     |
| tat | aac | ctt | gct | gac | ttt | ggt | aac  | aag | att | cta | ttt | ggt | ttg | atc | att | 720 |
| Tyr | Asn | Leu | Ala | Asp | Phe | Val | Asn  | Lys | Ile | Leu | Phe | Gly | Leu | Ile | Ile |     |
| 225 |     |     |     |     | 230 |     |      |     |     | 235 |     |     |     |     | 240 |     |
| tgg | aat | ggt | gct | ggt | aaa | gaa | tct  | tct | aat | gct |     |     |     |     |     | 753 |
| Trp | Asn | Val | Ala | Val | Lys | Glu | Ser  | Ser | Asn | Ala |     |     |     |     |     |     |
|     |     |     |     | 245 |     |     |      |     | 250 |     |     |     |     |     |     |     |

<210> 53  
 <211> 251  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 53

Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15

Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110



|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Leu | Ala | Ala | Cys | Thr | Asn | Val | Ala | Ala | Ser | Leu | Phe | Lys | Lys | Leu |
|     | 115 |     |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Leu | Ala | Gly | Ser | Leu | Val | Met | Leu | Gly | Ala | Gly | Phe | Ala | Gly | Glu | Ala |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Gly | Leu | Ala | Pro | Val | Trp | Pro | Ala | Phe | Ile | Ile | Gly | Met | Ala | Gly | Trp |
|     | 145 |     |     |     | 150 |     |     |     | 155 |     |     |     |     |     | 160 |
| Leu | Tyr | Met | Ile | Tyr | Glu | Leu | Tyr | Met | Gly | Glu | Gly | Lys | Ala | Ala | Val |
|     |     |     |     | 165 |     |     | 170 |     |     |     |     |     |     | 175 |     |
| Ser | Thr | Ala | Ser | Pro | Ala | Val | Asn | Ser | Ala | Tyr | Asn | Ala | Met | Met | Met |
|     |     |     |     |     |     |     | 180 |     | 185 |     |     |     | 190 |     |     |
| Ile | Ile | Val | Val | Gly | Trp | Ala | Ile | Tyr | Pro | Ala | Gly | Tyr | Ala | Ala | Gly |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Tyr | Leu | Met | Gly | Gly | Glu | Gly | Val | Tyr | Ala | Ser | Asn | Leu | Asn | Leu | Ile |
|     |     | 210 |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Tyr | Asn | Leu | Ala | Asp | Phe | Val | Asn | Lys | Ile | Leu | Phe | Gly | Leu | Ile | Ile |
|     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
| Trp | Asn | Val | Ala | Val | Lys | Glu | Ser | Ser | Ser | Asn | Ala |     |     |     |     |
|     |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     |     |

```

<210> 54
<211> 753
<212> DNA
<213> Naturally occurring gamma proteobacterium

<220>
<221> CDS
<222> (1)..(753)
<223> Proteorhodopsin variant from pcr clone PalB7; GenBank #AF349999

<400> 54
atg ggt aaa tta tta ctg ata tta ggt agt gct att gcg ctt cca tca 48
Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

tca ttc tgg ctg gtt acg gct ggt atg tta gcg gca act gta ttc ttt 144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg 240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

aga ggt gtt tgg ata gat act ggt gat aca cca aca gta ttt aga tat 288

```

|   |     |
|---|-----|
| Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr     |     |
| 85 90 95  |     |
| att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta     | 336 |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu     |     |
| 100 105 110   |     |
| att ctt gcc gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt     | 384 |
| Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu     |     |
| 115 120 125   |     |
| cta gct ggt tca ttg gta atg tta ggt gct gga tct gca ggc gaa gct     | 432 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Ser Ala Gly Glu Ala     |     |
| 130 135 140   |     |
| gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg     | 480 |
| Gly Leu Ala Pro Val Leu Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp |     |
| 145 150 155 160   |     |
| tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta     | 528 |
| Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Val         |     |
| 165 170 175   |     |
| agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg     | 576 |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met     |     |
| 180 185 190   |     |
| att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt     | 624 |
| Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly     |     |
| 195 200 205   |     |
| tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctc ata     | 672 |
| Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile     |     |
| 210 215 220   |     |

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720  
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753  
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 55  
 <211> 251  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 55

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Ser Ala Gly Glu Ala  
 130 135 140  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Tyr Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met  
 180 185 190  
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
 195 200 205  
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
 210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 56  
 <211> 753  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium

<220>  
 <221> CDS  
 <222> (1)..(753)  
 <223> Proteorhodopsin variant from pcr clone PalB6; GenBank # AF349998

<400> 56 48  
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15  
 ttt gct gct gct ggc gat cta gat ata agt gat act gtt ggt gtt 96  
 Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30  
 tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192  
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
 aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
 att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
 att ctt gct gct tgt aca aat gtt gct gct tca tta ttt aag aag ctt 384  
 Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
 gga tta gct cct gta tgg cct gct ttc att att ggt atg gct gga tgg 480  
 Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
 tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
 agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg gtg 576  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val  
 180 185 190

att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624  
 ile ile val val gly trp ala ile tyr pro ala gly tyr ala ala gly  
 195 200 205

tac cta atg ggt ggc gaa ggt gta tac gct tca aac cta aac ctt ata 672  
 tyr leu met gly gly glu gly val tyr ala ser asn leu asn leu ile  
 210 215 220

tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720  
 tyr asn leu ala asp phe val asn lys ile leu phe gly leu ile ile  
 225 230 235 240

tgg aat gtt gct gtt aaa gaa tct tct aat gct 753  
 trp asn val ala val lys glu ser ser asn ala  
 245 250

<210> 57  
 <211> 251  
 <212> PRT  
 <213> Naturally occurring gamma proteobacterium

<400> 57  
 Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45



Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60  
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
65 70 75 80  
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
85 90 95  
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
100 105 110  
Ile Leu Ala Ala Cys Thr Asn Val Ala Ala Ser Leu Phe Lys Lys Leu  
115 120 125  
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
130 135 140  
Gly Leu Ala Pro Val Trp Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
145 150 155 160  
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
165 170 175  
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Val

180

185

190

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250

<210> 58

<211> 753

<212> DNA

<213> Naturally occurring gamma proteobacteria

<220>

<221> CDS

<222> (1)..(753)

<223> Proteorhodopsin variant from pcr clone PalB8; GenBank #AF350000

<400> 58

atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

|   |     |
|---|-----|
| ttt gct gct gct ggt ggc gat cta gat ata agt gat act gttt ggt gtt<br>Phe Ala Ala 20 Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val<br>30                  | 96  |
| tca ttc tgg ctg gtt aca gct gct ggt atg tta gcg gca act gtg ttc ttt<br>Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe<br>35 40 45        | 144 |
| ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act<br>Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr<br>50 55 60        | 192 |
| gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg<br>Val Ser Gly Leu Ile Thr 70 Phe Trp His Tyr Leu Tyr Met<br>65 75 80                 | 240 |
| aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat<br>Arg Gly Val Trp 85 Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr<br>90 95        | 288 |
| att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta<br>Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu<br>100 105 110     | 336 |
| att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt<br>Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu<br>115 120 125     | 384 |
| cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct<br>Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala<br>130 135 140     | 432 |
| gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg<br>Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp<br>145 150 155 160 | 480 |

|   |     |
|---|-----|
| tta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta<br>Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Lys Ala Ala Val<br>165 170 175         | 528 |
| agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg<br>Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met<br>180 185 190     | 576 |
| att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt<br>Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly<br>195 200 205     | 624 |
| tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata<br>Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile<br>210 215 220     | 672 |
| tat aac ctt gct gac ctt gtt aac aag att cta ttt ggt ttg atc att<br>Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile<br>225 230 235 240 | 720 |
| tgg aat gtt gct gtt gtt aaa gaa tct tct aat gct<br>Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala<br>245 250   | 753 |
| <br><210> 59<br><211> 251<br><212> PRT<br><213> Naturally occurring gamma proteobacteria  |     |
| <br><400> 59  |     |
| Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser<br>1 5 10 15  |     |

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45  
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp

```

145          150          155          160
Leu Tyr Met Ile Tyr Glu Leu Tyr Tyr Met Gly Glu Gly Lys Ala Ala Val
165          170
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met
180          185          190
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly
195          200          205
Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile
210          215          220
Tyr Asn Leu Ala Asp Leu Val Asn Lys Ile Leu Phe Gly Leu Ile Ile
225          230          235          240
Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala
245          250
<210> 60
<211> 753
<212> DNA
<213> Naturally occurring gamma proteobacteria
<220>
<221> CDS
<222> (1) .. (753)

```

<223> Proteorhodopsin variant from pcr clone PalE1;GenBank# AF350001

```
<400> 60
atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca 48
Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser
1 5 10 15

ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96
Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val
20 25 30

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt 144
Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe
35 40 45

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192
Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr
50 55 60

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met
65 70 75 80

aga ggt gtt tgg ata gac act ggt gat acc cca aca gta ttc aga tat 288
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr
85 90 95

att gat tgg tta tta act gtt cca tta caa gtg gtt gag ttc tat cta 336
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu
100 105 110

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu
115 120 125
```

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
 gga tta gct cct gta tta cct gct ttc att att ggt atg gct gga tgg 480  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
 tta tac atg att tat gag cta tat atg ggt gaa ggc aag gct gct gta 528  
 Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
 agt act gca agt cct gct gtt aac cct gca tac aac gca atg atg atg 576  
 Ser Thr Ala Ser Pro Ala Val Asn Pro Ala Tyr Asn Ala Met Met Met  
 180 185 190  
 att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624  
 Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
 195 200 205  
 tac cta atg ggt ggc gaa ggt gta tac gct tca aac tta aac ctt ata 672  
 Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
 210 215 220  
 tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720  
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
 225 230 235 240  
 tgg aat gtt gct gtt aaa gaa tct tct aat gct 753  
 Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 61



<211> 251  
<212> PRT  
<213> Naturally occurring gamma proteobacteria

<400> 61

Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
50 55 60

Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
65 70 75 80

Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
85 90 95

Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Val Val Glu Phe Tyr Leu  
100 105 110

Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu

|   |     |     |
|---|-----|-----|
| 115   | 120 | 125 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala |     |     |
| 130   | 135 | 140 |
| Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp |     |     |
| 145   | 150 | 155 |
| Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val |     |     |
| 165   | 170 | 175 |
| Ser Thr Ala Ser Pro Ala Val Asn Pro Ala Tyr Asn Ala Met Met Met |     |     |
| 180   | 185 | 190 |
| Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly |     |     |
| 195   | 200 | 205 |
| Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile |     |     |
| 210   | 215 | 220 |
| Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile |     |     |
| 225   | 230 | 235 |
| Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala                     |     |     |
| 245   | 250 |     |

<210> 62  
 <211> 753  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium  
  
 <220>  
 <221> CDS  
 <222> (1)..(753)  
 <223> Proteorhodopsin variant from pcr clone Pale6; GenBank#AF350002  
  
 <400> 62 48  
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15  
  
 ttt gct gct gct ggt ggc gat cta gat ata agt gat act gtt ggt gtt 96  
 Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30  
  
 tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gta ttc ttt 144  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe  
 35 40 45  
  
 ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act 192  
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60  
  
 gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tac atg 240  
 Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
 65 70 75 80  
  
 aga ggt gtt tgg ata gat act ggt gat aca cca gta ttt aga tat 288  
 Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
 85 90 95

att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336  
 Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
 100 105 110  
  
 att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384  
 Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
 115 120 125  
  
 cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432  
 Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
 130 135 140  
  
 ggt tta gct cct gta tta cct gct gct ttc att att ggt atg gct gga tgg 480  
 Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
 145 150 155 160  
  
 tta tac atg att tat gag cta cat atg ggt gaa ggt aag gct gct gta 528  
 Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Gly Lys Ala Ala Val  
 165 170 175  
  
 agt act gca agt cct gct gct gtt aac tct gca tac aac gca atg atg aag 576  
 Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys  
 180 185 190  
  
 att att gtt att gga tgg gca att tat cct gct gga tat gct gct ggt 624  
 Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
 195 200 205  
  
 tac cta atg agt ggt gac ggt gta tac gct tca aac tta aac ctt ata 672  
 Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
 210 215 220  
  
 tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att 720  
 Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile

|   |     |     |     |     |
|---|-----|-----|-----|-----|
| 225   | 230 | 235 | 240 |     |
| tgg aat gtt gct gtt aaa gaa tct tct aat gct                     |     |     |     | 753 |
| Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala                     |     |     |     |     |
| 245   | 250 |     |     |     |
| <210> 63  |     |     |     |     |
| <211> 251   |     |     |     |     |
| <212> PRT   |     |     |     |     |
| <213> Naturally occurring gamma proteobacterium                 |     |     |     |     |
| <400> 63  |     |     |     |     |
| Met Gly Lys Leu Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser |     |     |     |     |
| 1   | 5   | 10  | 15  |     |
| Phe Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val     |     |     |     |     |
| 20  | 25  | 30  |     |     |
| Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Thr Val Phe Phe     |     |     |     |     |
| 35  | 40  | 45  |     |     |
| Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr |     |     |     |     |
| 50  | 55  | 60  |     |     |
| Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met |     |     |     |     |
| 65  | 70  | 75  | 80  |     |
| Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr |     |     |     |     |

|   |     |     |     |
|---|-----|-----|-----|
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu | 85  | 90  | 95  |
| 100   | 105 | 110 |     |
| Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu | 115 | 120 | 125 |
|   |     |     |     |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala | 130 | 135 | 140 |
|   |     |     |     |
| Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp | 145 | 150 | 155 |
|   |     |     | 160 |
| Leu Tyr Met Ile Tyr Glu Leu His Met Gly Glu Lys Ala Ala Val     | 165 | 170 | 175 |
|   |     |     |     |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Lys | 180 | 185 | 190 |
|   |     |     |     |
| Ile Ile Val Ile Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly | 195 | 200 | 205 |
|   |     |     |     |
| Tyr Leu Met Ser Gly Asp Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile | 210 | 215 | 220 |
|   |     |     |     |

09847513.080801

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
 225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
 245 250

<210> 64  
 <211> 753  
 <212> DNA  
 <213> Naturally occurring gamma proteobacterium

<220>  
 <221> CDS  
 <222> (1)..(753)  
 <223> Proteorhodopsin variant from pcr clone Pale7; GenBank# AF350003

<400> 64  
 atg ggt aaa tta tta ctg ata tta ggt agt gct att gca ctt cca tca  
 Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
 1 5 10 15 48

ttt gct gct gct ggc gat cta gat ata agt gat act gtt ggt gtt  
 Phe Ala Ala Ala Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
 20 25 30 96

tca ttc tgg ctg gtt aca gct ggt atg tta gcg gca act gtg ttc ttt  
 Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
 35 40 45 144

ttt gta gaa aga gac caa gtc agc gct aag tgg aaa act tca ctt act  
 Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr  
 50 55 60 192

gta tct ggt tta att act ggt ata gct ttt tgg cat tat ctc tat atg 240  
Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met  
65 70 75 80

aga ggt gtt tgg ata gat act ggt gat acc cca aca gta ttc aga tat 288  
Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr  
85 90 95

att gat tgg tta tta act gtt cca tta caa atg gtt gag ttc tat cta 336  
Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu  
100 105 110

att ctt gct gct tgt aca agt gtt gct gct tca tta ttt aag aag ctt 384  
Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu  
115 120 125

cta gct ggt tca tta gta atg tta ggt gct gga ttt gca ggc gaa gct 432  
Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala  
130 135 140

gga tta gct cct gta tta cct gct gct ttc att att ggt atg gct gga tgg 480  
Gly Leu Ala Pro Val Leu Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp  
145 150 155 160

cta tac atg att tat gag cta tat atg ggt gaa ggt aag gct gct gta 528  
Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val  
165 170 175

agt act gca agt cct gct gtt aac tct gca tac aac gca atg atg atg 576  
Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met  
180 185 190

att att gtt gtt gga tgg gca att tat cct gct gga tat gct gct ggt 624  
Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly



|   |     |     |     |
|---|-----|-----|-----|
| 195   | 200 | 205 |     |
| tac cta atg ggt ggc gaa ggc gta tac gct tca aac tta aac ctt ata |     |     | 672 |
| Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile |     |     |     |
| 210   | 215 | 220 |     |
| tat aac ctt gct gac ttt gtt aac aag att cta ttt ggt ttg atc att |     |     | 720 |
| Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile     |     |     |     |
| 225   | 230 | 235 | 240 |
| tgg aat gtt gct gtt aaa gaa gaa tct tct aat gct                 |     |     | 753 |
| Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala                     |     |     |     |
|   | 245 | 250 |     |

<210> 65  
<211> 251  
<212> PRT  
<213> Naturally occurring gamma proteobacterium

$<400>$  65

Met Gly Lys Leu Leu Ile Leu Gly Ser Ala Ile Ala Leu Pro Ser  
1 5 10 15

Phe Ala Ala Ala Gly Gly Asp Leu Asp Ile Ser Asp Thr Val Gly Val  
20 25 30

Ser Phe Trp Leu Val Thr Ala Gly Met Leu Ala Ala Thr Val Phe Phe  
35 40 45

Phe Val Glu Arg Asp Gln Val Ser Ala Lys Trp Lys Thr Ser Leu Thr

|   |     |    |     |
|---|-----|----|-----|
| 50  | 55  | 60 |     |
| Val Ser Gly Leu Ile Thr Gly Ile Ala Phe Trp His Tyr Leu Tyr Met |     |    |     |
| 65  | 70  | 75 | 80  |
| Arg Gly Val Trp Ile Asp Thr Gly Asp Thr Pro Thr Val Phe Arg Tyr |     |    |     |
| 85  | 90  |    | 95  |
| Ile Asp Trp Leu Leu Thr Val Pro Leu Gln Met Val Glu Phe Tyr Leu |     |    |     |
| 100   | 105 |    | 110 |
| Ile Leu Ala Ala Cys Thr Ser Val Ala Ala Ser Leu Phe Lys Lys Leu |     |    |     |
| 115   | 120 |    | 125 |
| Leu Ala Gly Ser Leu Val Met Leu Gly Ala Gly Phe Ala Gly Glu Ala |     |    |     |
| 130   | 135 |    | 140 |
| Gly Leu Ala Pro Val Leu Pro Ala Phe Ile Ile Gly Met Ala Gly Trp |     |    |     |
| 145   | 150 |    | 155 |
| Leu Tyr Met Ile Tyr Glu Leu Tyr Met Gly Glu Gly Lys Ala Ala Val |     |    |     |
| 165   | 170 |    | 175 |
| Ser Thr Ala Ser Pro Ala Val Asn Ser Ala Tyr Asn Ala Met Met Met |     |    |     |
| 180   | 185 |    | 190 |

Ile Ile Val Val Gly Trp Ala Ile Tyr Pro Ala Gly Tyr Ala Ala Gly  
195 200 205

Tyr Leu Met Gly Gly Glu Gly Val Tyr Ala Ser Asn Leu Asn Leu Ile  
210 215 220

Tyr Asn Leu Ala Asp Phe Val Asn Lys Ile Leu Phe Gly Leu Ile Ile  
225 230 235 240

Trp Asn Val Ala Val Lys Glu Ser Ser Asn Ala  
245 250